



„Systematische Spielbeobachtung im internationalen Leistungsfußball“

Kumulative Dissertation

zur Erlangung des Doktorgrades
der Kulturwissenschaftlichen Fakultät
der Universität Bayreuth

vorgelegt von

Tianbiao Liu

am

29.01.2014

Inhalt

- 1 Zusammenfassung der schriftlichen Promotionsleistung**
- 2 Einordnung der kumulativen Dissertation in den Forschungskontext**
- 3 Verzeichnis wissenschaftlicher Veröffentlichungen**
- 4 Wissenschaftliche Kongresse**
- 5 Lehrveranstaltungen**
- 6 Betreute Abschlussarbeiten und Seminararbeiten**
- 7 Erklärung**
- 8 Lebenslauf**
- 9 Veröffentlichungen**

1 Zusammenfassung der schriftlichen Promotionsleistung

Im Rahmen der Zielvereinbarung zwischen dem Doktoranden und dem Fachmentorat wurde eine schriftliche Promotionsleistung, bestehend aus drei wissenschaftlichen Arbeiten (kumulative Promotion) zum Themenbereich „Sportspielforschung“ festgeschrieben (siehe Anlage). Zum Zwecke der Promotion werden im Einzelnen vorgelegt:

Zeitschriftenartikel (Peer reviewed)

- (1) Liu, T., Hohmann, A., Castro, C. & Sörgel, V. (2014 angenommen, zur Veröffentlichung vorgesehen in 2015, s. Bescheinigung). Systematische Spielanalyse im Frauenfußball – Finale des Algarve Cups 2012 Deutschland – Japan. *Zeitschrift für Angewandte Trainingswissenschaft*.
- (2) Liu, T. & Hohmann, A. (2013). Applying Data Mining to Analyze the Different Styles of Offense between Manchester United and FC Barcelona in the European Champions League Final. *International Journal of Sports Science and Engineering*, 7(02), 067-078.
- (3) Liu, T. & Hohmann, A. (2013). Applying the Markov Chain theory to Analyze the Attacking Actions between FC Barcelona and Manchester United in the European Champions League final. *International Journal of Sports Science and Engineering*, 7(02), 079-086.

Die zentralen Ergebnisse der eingereichten Arbeiten werden im Folgenden zusammengefasst. Mit Verweis auf die vorgenannten Einzelpublikationen wird auf ein abschließendes Literaturverzeichnis verzichtet.

2 Einordnung der kumulativen Dissertation in den Forschungskontext

Die Sportspielbeobachtung gehört zu den wichtigen Aufgaben der trainingswissenschaftlichen Leistungsdiagnostik. Im Bereich der leistungsdiagnostisch orientierten Spielanalyse im Fußball ist vor allem das Verfahren der Systematischen Spielerbeobachtung bekannt (Loy, 1995; Hohmann, & Rommel, 1994). Im Rahmen der Fußballspielanalyse entwickelte sich dabei das analytische Verfahren über drei Entwicklungsphasen (a) „Traditionelle deskriptive Betrachtung des Spielverhaltens“, (b) „Analytische Betrachtung des Spielverhaltens“ und (c) „Simulative Betrachtung des Spielverhaltens“. Die *Traditionelle deskriptive Betrachtung des Spielverhaltens*, als ein grundlegendes Verfahren zur Analyse des Fußballspiels, findet seit Beginn der wissenschaftlichen Beschäftigung mit der Sportart Fußball Anwendung. Im Rahmen der *Simulativen Betrachtung des Spielverhaltens* werden geeignete mathematische Modelle (z.B. Markovkette) einbezogen, mit denen das Fußballspiel adäquat simuliert werden kann. Die *Analytische Betrachtung des Spielverhaltens* beschäftigt sich ebenfalls mit dem Konstruieren von mathematischen Modellen (z.B. durch einen Apriori-Algorithmus), mit dessen Hilfe sich ein Fußballspiel aus einer ganzheitlichen und verknüpfenden Perspektive beschreiben lässt. Darüber hinaus lassen sich durch verbesserte Algorithmen zudem weitere diagnostische Ergebnisse über ein Fußballspiel ableiten.

Nachdem lange Zeit die Anwendung der traditionellen deskriptiven Analyse im Fußball und anderen Sportarten Anwendung fand, ziehen Trainingswissenschaftler heute neuere Ansätze im Sport heran. Diese neuen Ansätze stammen ursprünglich aus der Mathematik, die auf Sportspiele

übertragen werden. Mit der Einführung des *Mathematisch - simulativen Ansatzes* durch Lames (1991), bei dem Sportspiele mittels der Markov-Ketten-Theorie analysiert werden, wurde eine neue forschungsmethodische Debatte in der Sportspielforschung angeregt. Die Markov-Ketten-Methode stützt sich auf eine Übergangsmatrix, die alle Spielzustände beinhaltet und der Berechnung der Leistungsrelevanz dient. Dadurch lässt sich schließlich ein wichtiger Spieleinflussfaktor identifizieren. Die Umsetzung und Ergebnisse dieser Theorie demonstrierte Lames (1991) beispielhaft in der Sportart Tennis, bevor sie danach von Zhang (2003) auch in der Sportart Tischtennis angewendet wurde. In den letzten Jahren wurden sowohl die videobasierte Sportspielbeobachtung als auch die Analysetheorie weiterentwickelt, wobei die Computertechnologie sowie die Datenbanktechnologie einen entscheidenden Einfluss hatten. Mit den sogenannten Data-Mining-Verfahren im Sport können überzufällige Spieltendenzen und Verbindungen zwischen verschiedenen Spielzuständen herausgefunden werden. Auf Basis des *Association Rule Algorithmus* und anderen verbesserten Algorithmen innerhalb des Data-Mining haben beispielsweise Yu, Zhang und andere Wissenschaftler (2008) einen entscheidenden Beitrag zum Erfolg der chinesischen Tischtennis-Mannschaft bei den Olympischen Spielen 2008 in Peking geleistet.

Anknüpfend an diese kurz skizzierte Entwicklung wurden von mir drei Beiträge zum Themenbereich „Sportspielforschung“ veröffentlicht. Dabei wurde mittels eines mathematisch-deskriptiven und einem mathematisch-simulativ Ansatzes sowie einer empirisch-analytischen Betrachtung des Spielverhaltens das komplexe Mannschaftssportspiel Fußball analysiert. Auf der Basis der drei Ansätze behandeln die drei Beiträge europäische Champions League Finale 2011 (Liu & Hohmann, 2013; Liu & Hohmann, 2013) sowie das Algarve Cup Finale 2012 (Liu & Hohmann, 2014 angenommen).

2.1 Untersuchungsmethode

2.1.1 Mathematisch-simulative Methode (Markov-Kette)

In den Beiträgen zur Spielbeobachtung erfolgte das forschungsmethodische Vorgehen in Anlehnung an den mathematisch-simulativen Ansatz. Dieser umfasst gemäß Pfeiffer (2010) folgende Arbeitsschritte:

- „(1) Strukturierung des Sportspiels in Zustände (Entwicklung eines Zustandsmodells),
- (2) Konstruktion des Beobachtungsmodells,
- (3) Mathematische Modellierung mit Markov-Ketten,
- (4) Definition der zu untersuchenden taktischen Verhaltensweisen, d.h. der Zustandsübergänge,
- (5) Simulation zur Bestimmung der Leistungsrelevanz der definierten taktischen Verhaltensweisen und
- (6) Priorisierung der taktischen Verhaltensweisen.

- (1) Es wird ein Kategoriensystem über alle Spielsituationen des jeweiligen Sportspiels erarbeitet, welches die Zustände enthält, in denen sich das Spiel befinden kann (Zustandsmodell). Somit wird das Sportspiel als System aufgefasst, das sich schrittweise durch eine Menge von

Zuständen bewegt. Mit der Modellierung auf der Zustandsebene werden taktische Verhaltensweisen in den Zustandsübergängen abgebildet (Zustand-Übergang-Modell). Das leistungsdiagnostische Erkenntnisinteresse setzt voraus, dass ein Leistungskriterium modellseitig implementiert wird. Aus leistungsdiagnostischer Perspektive wird das Erreichen des Spielzieles als Kriterium der Leistung im Sportspiel definiert, welches in Form von terminalen Zuständen in das Modell aufzunehmen ist. Für die Fußballspiele werden, auf Grund der wenigen Attacken in jedem Fußballspiel, „das Eindringen in den 35-m-Raum vor dem gegnerischen Tor“ als zweckdienlich angesehen.

- (2) Es gilt ein Beobachtungssystem zu erarbeiten, in dem die Zustände operational definiert sind. Das Beobachtungssystem muss hinsichtlich seiner Objektivität durch die Quantifizierung der Beobachterübereinstimmung (Interrater- Reliabilität) geprüft werden.
- (3) Mit einem Zustands-Übergangs-Modell kann das Spielverhalten über die Übergangswahrscheinlichkeiten zwischen den einzelnen Zuständen abgebildet werden. Dabei kann die Übergangsfunktion als Markov-Kette aufgefasst werden, wenn man folgende Eigenschaften akzeptiert: (1.) Die Übergangswahrscheinlichkeit zwischen zwei Zuständen hängt nur vom vorausgehenden Zustand ab (Markov Eigenschaft) und (2.) die Übergangswahrscheinlichkeit zwischen zwei Zuständen ist unabhängig von der Position in der Ereignisfolge (Ketten Eigenschaft). Markov-Ketten 1. Ordnung lassen sich in einer zweidimensionalen Übergangsmatrix darstellen. In dieser Übergangsmatrix werden die prozentualen Häufigkeiten der Übergänge von einem Zustand in einen Folge-Zustand festgehalten. Die Übergänge von einem in den anderen Zustand sind Ausdruck des taktischen Verhaltens der Mannschaft. Weiterhin werden in der Theorie der Markov-Ketten verschiedene Typen von Zuständen unterschieden. Ein Zustand heißt absorbierend, wenn er nicht mehr verlassen werden kann, d.h. der Prozess endet in diesem Zustand. Alle anderen nicht-absorbierenden Zustände werden „innere“ Zustände genannt. Als absorbierende Zustände werden die terminalen Zustände (z.B. „into 35 m“, s.o.) definiert. Die Übergangswahrscheinlichkeit in diese terminalen Zustände wird als Erfolgswahrscheinlichkeit („into 35m“) bezeichnet und stellt die leistungsdiagnostisch relevante Größe dar. Die Erfolgswahrscheinlichkeit der Mannschaft für das Erreichen des Spielziels lässt sich mit dem stochastischen Modell der Markov-Kette aus einer empirischen Übergangsmatrix iterativ durch Matrizenmultiplikation berechnen.
- (4) Aus den Zustandsübergängen sind diejenigen auszuwählen, die im Kontext der jeweiligen leistungsdiagnostischen Zielstellung relevante taktische Verhaltensweisen repräsentieren. Sie sind bezüglich ihres semantischen Inhalts zu ordnen, weshalb eine gewisse Parallelität zum leistungsdiagnostischen Arbeitsschritt *Ordnung interner Beziehungen* (Letzelter, 1978) besteht. Die Ordnung ist hier jedoch theoretisch-fachwissenschaftlicher und nicht empirisch-statistischer Natur.
- (5) Die Übergangswahrscheinlichkeiten werden numerisch so manipuliert, dass damit eine jeweils höhere oder geringere Häufigkeit der taktischen Verhaltensweise dargestellt wird. Für die im Sinne der taktischen Verhaltensweise manipulierte Übergangsmatrix werden anschließend erneut die Erfolgswahrscheinlichkeiten (nach den oben genannten Verfahren) berechnet. Das Maß der Leistungsrelevanz einer taktischen Verhaltensweise (δAP) ist dann

die Differenz der Erfolgswahrscheinlichkeiten, die sich gegenüber der ursprünglichen empirischen Ausgangs-Übergangsmatrix ergibt.

- (6) Die simulativ ermittelten Leistungsrelevanzen der Verhaltensweise (δAP) liegen als quantitatives Maß vor, sind untereinander vergleichbar und können im Sinne der Priorisierung direkt in eine Rangfolge gebracht werden. In Abhängigkeit vom Zustandsmodell ist die Priorisierung im Hinblick auf die Strukturierung des Verhaltensstroms vorzunehmen, d.h. es sollten lediglich konkurrierende taktische Verhaltensweisen miteinander verglichen werden.“ (Pfeiffer, 2010, S. 5-6)

2.1.2 Mathematisch-analytische Methode (Apriori Algorithmus)

Parallel zur mathematisch-simulativen Methode wurde im Rahmen der Dissertation mit Hilfe des Data-Minings der mathematisch-analytische Ansatz angewendet, wobei hier der Spielfluss in einem Fußballspiel als Ganzes behandelt wird. Wörtlich übersetzt bedeutet Data Mining „Graben in großen Datenmengen“, wobei das Ziel verfolgt wird, neues und schwer zugängliches Wissen zu entdecken (Alpar, Niedereichholz, Bibel & Kruse, 2000). Data Mining kann somit als Prozess der Gewinnung von neuem und nützlichem Wissen aus großen Datenmengen unter Anwendung algorithmischer Methoden definiert werden (Tamtam, 2012, S. 2).

In Fußballspielen werden nach El Jerroudi (2001) beim Data-Mining häufig das Association Rule-Mining und der entsprechende Apriori Algorithmus sowie dessen modifizierte Versionen verwendet. Grundsätzlich gilt dabei $I = \{i_1, i_2, \dots, i_n\}$ eine Menge von Objekten (Items). D sei eine Menge von Transaktionen, wobei jede Transaktion eine Menge von Objekten ist, also $T \subseteq I$. Mit jeder Transaktion ist ein eindeutiger Bezeichner, genannt TID assoziiert. Eine Assoziationsregel ist eine Implikation der Form $X \Rightarrow Y$, wobei X und Y Untermengen von I sind (X und Y haben keine gemeinsamen Elemente). Eine Regel $X \Rightarrow Y$ hat den *Konfidenzwert* c , falls $c\%$ der Transaktionen aus D , die X enthalten, auch Y enthalten. Eine Regel $X \Rightarrow Y$ hat den *Support* s , wenn $s\%$ der Transaktionen aus D , X vereinigt und Y enthalten. Eine Regel $X \Rightarrow Y$ hat den *Wichtigkeitswert* (*Importance*), der in vielen Fällen die Qualität der Regel zeigt.

Normalerweise besteht der Data-mining Prozess aus fünf Arbeitsschritten, auf die gemäß Tamtam (2012) im Folgenden eingegangen wird: „(1) Datenauswahl, (2) Datenvorbereitung, (3) Datentransformation, (4) Data Mining, (5) Interpretation und Evaluierung.

(1) *Datenauswahl*: Die relevanten Daten werden für die Analyse von riesigen Datenbeständen selektiert. [Im Fußballspiel werden während der Videobeobachtung alle technischen und taktischen Aktionen beider Mannschaften mit Hilfe der Indexnummern (Kategoriensystem) in einer Tabelle notiert. Zur Identifikation der einzelnen Spieler dient die Rückennummer; d.V]

(2) *Datenvorbereitung*: Dabei werden fehlende Werte behandelt (z.B. durch Weglassen der entsprechenden Datensätze oder Ersetzen der fehlenden Werte durch Standardwerte), Ausreißer identifiziert bzw. entfernt und Inkonsistenzen in den Daten gelöscht.

(3) *Datentransformation*: Hier ist die Hauptaufgabe die Daten so umzuwandeln, dass sie für das entsprechende Data-Mining-Verfahren geeignet sind. Es geht u.A. darum, neue Attribute durch

Berechnung oder Aggregation zu generieren oder Merkmalsausprägungen einiger Attribute zu normalisieren.

(4) *Data Mining*: In diesem Schritt wird die eigentliche Analyse der Daten durchgeführt. Auf der Basis des vorbereiteten Datenbestandes werden algorithmische Methoden eingesetzt, um unter der Oberfläche liegende Gesetzmäßigkeiten und Muster zu extrahieren. Schließlich werden die Konfidenz, der Support und/oder die Wichtigkeit berechnet.

(5) *Interpretation und Evaluation*: Die entdeckten Gesetzmäßigkeiten müssen letztendlich mit Hilfe von Visualisierungstechniken präsentiert werden, um sie leichter interpretieren bzw. beurteilen zu können. Falls sie einen hohen Grad an Signifikanz aufweisen, können sie für die Lösung von sportspezifischen Problemstellungen eingesetzt werden.“(Tamtam, 2012, S. 2-3)

2.2 Untersuchungsergebnisse

Mit dem Beitrag „**Applying the Markov Chain theory to Analyze the Attacking Actions between FC Barcelona and Manchester United in the European Champions League final**“ (Liu & Hohmann, 2013) wurde der mathematisch-simulative Ansatz der Leistungsdiagnostik in die Sportart Fußball eingeführt. Im Fokus der Untersuchung stand die Frage, welche **Spielkombinationen** eine wichtige Rolle in Bezug auf die Torchancen (into 35m) spielen.

Die Fragestellung wurde unter Verwendung des Spielerbeobachtungsmodells bearbeitet, um darin die einzelnen Merkmals-Spielzustände zu charakterisieren. Das Spielerbeobachtungsmodell bildet jeden einzelnen Ballpass ab. Als Spielziel wird in dem Spieler- Beobachtungsmodell der Zustand „into 35m“ (Torchance, absorbierender Zustand) abgegrenzt.

Die Spieldaten wurden im Champions League-Finale 2011 gesammelt, das qualitätsmäßig zum Topklassenniveau in der Welt gehört. Alle Spielaktionen wurden videografisch erfasst, notiert und analysiert. Dabei wurden über 1.200 Spieldaten aufgezeichnet. Die *Objektivitätsprüfung* ergab für die Beobachtungsmodelle exzellente Interrater-Reliabilitäten (Cohen's Kappa-Wert) sind $\kappa = 0.766$ für die Einzelspieler und $\kappa = 0.625$ für die Zonen. Die Höhe der Kappa-Werte reflektiert die Schwierigkeit, alle Daten aus dem Video manuell zu notieren. Die Modellvalidierung im Sinne eines Vergleichs zwischen Modell- und Originalverhalten erfolgte korrelationsanalytisch über die Prüfung des Zusammenhangs zwischen den beobachteten und den theoretisch ermittelten Erfolgswahrscheinlichkeiten. Mit dem Korrelationskoeffizienten $r = 0.96$ konnten schließlich sehr zufrieden stellende Anpassungen des mathematischen Modells (Markov-Ketten) an die empirischen Daten erreicht werden.

Im Ergebnis konnte gezeigt werden, dass die Top-Spieler eine wesentliche Rolle gespielt haben. Insbesondere in der zweiten Halbzeit fallen z.B. Messi (FC Barcelona) und Rooney (Manchester United) auf, die relativ hohe Werte der leistungsrelevanten der Verhaltensweise (δAP) aufweisen. Anhand dieser quantitativen Kenngrößen des Leistungseinflusses können die taktischen Verhaltensweisen in eine Rangfolge gebracht werden (Priorisierung), anhand derer sich konkrete Handlungsempfehlungen für den Spiel ableiten lassen. Daraus folgt, dass eine Mannschaft die positiven Spielkombinationen verstärken und die Spielfehler von negativen Spielpässen

reduzieren muss. Die Unterschiedsprüfung bezüglich der Spielhalbzeiten ergab folgende Ergebnisse.

- In der ersten Halbzeit stellen für Manchester United, "Patrice Evra – Park Ji-Sung"; "Michael Carrick - Javier Hernandez" und "Patrice Evra - Javier Hernandez", sowie "Giggs – Park Ji-Sung" positive und effektive Spielkombinationen dar. Negativ hingegen sind "Patrice Evra – Michael Carrick"; "Javier Hernandez – Michael Carrick" und "Patrice Evra - Vidic".
- Für den FC Barcelona haben "Xavi Hernandez – Eric Abidal", "David Villa – Xavi Hernandez" und "Andres Iniesta – Eric Abidal" zu Torchance „into 35m“ beigetragen.
- Negative Spielkombinationen sind "Daniel Alves - David Villa", "Xavi Hernandez - David Villa", "Messi - David Villa".
- Entsprechend sind in der 2. Halbzeit für Manchester United, "Van Der Sar - Rooney", "Van der Sar - Hernandez", "Rooney - Hernandez" and "Carrick - Hernandez" positiv und "Patrice Evra – Michael Carrick"; "Javier Hernandez – Michael Carrick" und "Patrice Evra - Vidic" negativ.
- Für den FC Barcelona stellen in der zweiten Halbzeit, "Daniel Alves - Messi", "Xavi Hernandez - Messi", "Sergio Busquets - Messi" und "David Villa – Messi" positive Kombinationen dar. Negative Kombinationen sind "Messi - Keita", "Messi - Xavi Hernandez", "Sergio Busquets - Keita".

Unter dem Titel „**Applying Data Mining to Analyze the the Different Styles of Offense between Manchester United and FC Barcelona in the European Champions League Final**“ (Liu & Hohmann, 2013) habe ich außerdem die neue Data-Mining-Methode auf das Fußballspiel angewandt. Mit diesem mathematisch-analytischen Ansatz können typische Spieltendenzen und Verbindungen zwischen verschiedenen Spielern im Fußballspiel herausgefunden werden. Dies stellt somit eine Weiterentwicklung der Mathematisch-simulativen Methode (Markov-Ketten) (Lames, 1991) dar. Die Markov-Kette zeigt die direkten positiven oder negativen Verbindungen im Spiel, wobei der Apriori Algorithmus (Agrawal & Srikant, 1994) aus dem gleichen Spiel zusätzlich indirekte Verbindungen darstellen kann. Die Ergebnisse beschreiben die Ballauftendenz, wie z.B. der Ball in ein paar Schritten von Spieler A zu Spieler B gepasst wird.

Um die Spieltendenz zu beschreiben, wurden zwei alternative Beobachtungsmodelle für die Spielanalyse entwickelt. Zum einen handelt es sich dabei um das (1) Spielermodell und zum anderen um das (2) Zonemodell. Mit „Microsoft SQL Server data mining add-in for Excel“ wurden *Wahrscheinlichkeits-* und *Wichtigkeits-Werte* (Probability und Importance Werten) ausgerechnet, wodurch sich wichtige Association Rules, wie z.B. „5 = Existing, 1 = Existing -> 15 = Existing“, deren Probability 83% und Importance 0.63 ist, ermitteln lassen. Diese Rules (Regel) bedeutet, dass der Qualitätswert für diese Regel 0.63 beträgt, wenn Spieler Nr. 5 und Nr. 1 in einem Spielzug zusammen vorkommen und gleichzeitig Spieler Nr. 15 mit einer 83% Wahrscheinlichkeit in diesem Spielzug ebenfalls auftaucht.

Um diese Ergebnisse besser zu verstehen, können alle Association Rules und Spielmuster (Verhalten) als eine Grafik beider Mannschaften abgebildet werden. Aus dieser Spieltendenzgrafik geht hervor, dass im gesamten Spiel des FC Barcelona viel von den

Mittelfeldspielern abhing. In der ersten Halbzeit wurde das Mittelfeld von Manchester United besonders im Angriff auch Barcelona beschränkt. Aus diesem Grund modifizierte Manchester United die Spielstrategie in der zweiten Halbzeit, worauf die Pässertendenz zum Spielmacher Rooney hindeutet. Allerdings ist ersichtlich, dass sich die Situation in der zweiten Halbzeit nicht signifikant änderte und der FC Barcelona das Mittelfeld auch weiterhin kontrollierte.

Nach dem Versuch zwei separate Methoden in den beiden vorgestellten Artikeln anzuwenden, um das Spiel systematisch beobachten und analysieren zu können, zielte es eine weitere Forschungsarbeit darauf ab, ob eine Analyse für die Nationalmannschaft zu erstellen. Im Beitrag **„Systematische Spielanalyse im Frauenfußball – Finale des Algarve Cups 2012 Deutschland – Japan“** (Liu, Castro, Sörgel & Hohmann, 2014) wurde das Finale des Algarve-Cups 2012 mit der „Traditionellen deskriptiven Betrachtung des Spielverhaltens“, der „Simulativen Betrachtung des Spielverhaltens“ und der „Analytischen Betrachtung des Spielverhaltens“ parallel analysiert. Alle drei Methoden ergänzen einander und schaffen daher ein gutes Methodengefüge, um das Spiel vielseitig auszuwerten.

Das Datenmaterial stammt aus Videoaufzeichnungen des Deutschen Fußball-Bundes (DFB). Unter Verwendung des Apriori- (Agrawal & Srikant, 1994) und des Markov-Algorithmus (Lames, 1991), des Importance-Konzeptes (Qualität der Association Rules zu den Spielkombinationen), Probability (Regelmäßigkeit der Association Rules zu den Spielkombinationen), und Erfolgswahrscheinlichkeit (Effizienz des Passspiels) wurden die Daten mit verschiedenen Computerprogrammen (Datenanalyse mittels Microsoft SQL Server Data Mining Add-in® und SSF zur Programmierung der Markov-Ketten) ausgewertet.

In die Untersuchung wurden von beiden Mannschaften insgesamt $n=1400$ Einzelaktionen einbezogen. Die *Objektivität* der systematischen Spielbeobachtung wurde mit Hilfe von Cohen's Kappa überprüft. Der Mittelwert für die Beobachtung der ballbesitzenden (1) Spielerinnen lag bei $k = 0.53$. Für die angespielten (2) Spielfeldzonen lag der Mittelwert bei $k = 0.60$. Die *Validität* des Spielbeobachtungsmodells wurde über den Vergleich der empirisch beobachteten und der theoretisch berechneten Werte ermittelt. Die Korrelation betrug $r = 0.96$ ($p < 0.01$; $n = 8$) und belegt die gute Übereinstimmung der simulativ berechneten Ergebnisse mit den tatsächlich beobachteten Werten.

Im Ergebnis ließen sich vor allem für die gesamte Mannschaft typische kollektive Angriffskonzepte identifizieren. Der Vergleich des **Angriffsindex** basierte auf einer Statistik, nach der Deutschland das Spiel in der ersten Halbzeit überwiegend kontrollierte. Zudem unterstrich auch das Halbeitergebnis von 2:1 Toren die deutsche Dominanz. In der zweiten Halbzeit hingegen übernahm Japan zunehmend die Kontrolle des Spielgeschehens und dominierte den Spielverlauf über längere Zeit, auch wenn Japan das Spiel schlussendlich verlor. Die Differenz der Werte des Angriffsindex zwischen Deutschland und Japan in der zweiten Halbzeit ist statistisch signifikant ($t = -2,30$; $p < 0.05$). Aus den Importance- und Probabilitywerten (Wahrscheinlichkeit) wurden wiederum (1) Zonenmodelle (Inter-Zonen-Pässe) und (2) Spielermodelle

(Inter-Spieler-Pässe) aufgestellt. Das *Zonenmodell* zeigt, dass die deutsche Frauen-Nationalmannschaft in der ersten Halbzeit überwiegend auf der linken Seite, in der zweiten Halbzeit jedoch mehr auf der rechten Seite spielte. Das *Spielermodell* zeigt, dass in der ersten Halbzeit die Spielkombinationen zwischen den Spielerinnen Nr. 7, Nr. 4 und Nr. 5 sowie auch Spielzüge zwischen Nr. 11, Nr. 14 und Nr. 13 häufig in einer Spielsequenz erschienen sind. In der 2. Halbzeit traten die Spielkombinationen zwischen Nr. 13, Nr. 2 und Nr. 9 relativ häufig auf (hohe Probability). Diese hatten jedoch eine eher geringe Wichtigkeit (niedrige Importance bzw. niedrige Qualität der Association Rule). Die Ergebnisse der Markov-Ketten zeigen, dass in der ersten Halbzeit die Passspiel-Paarungen Nr. 2→Nr. 13, Nr. 4→Nr.13, Nr.2→Nr. 14 effizient waren. In der zweiten Halbzeit waren dies indes Nr. 3→Nr. 13, Nr. 2→Nr. 13, Nr. 2→Nr. 9.

2.3 Literaturverzeichnis

- Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In *Proc. 20th Int. Conf. Very Large Data Bases, VLDB* (Vol. 1215, pp. 487-499).
- Alpar, P., Niedereichholz, J., Bibel, W., & Kruse, R. (Eds.). (2000). *Data Mining im praktischen Einsatz*. Vieweg & Teubner Verlag.
- El Jerroudi, Z. (2001). Apriori-Algorithmus zur Entdeckung von Assoziationsregeln. Dortmund: Lehrstuhl für Künstliche Intelligenz der Universität Dortmund.
(<http://www-ai.cs.uni-dortmund.de/LEHRE/PG/PG402/seminar/Apriori.ppt>)
- Han, J., & Kamber, M. (2006). *Data mining: Concepts and Techniques*. Singapore: Elsevier
- Hohmann, A., & Rommel, G. (1994). Spielbeobachtung im Fußball. *Leistungssport*, 22(6), 41-46.
- Lames, M. (1991). *Leistungsdiagnostik durch Computersimulation: Ein Beitrag zur Theorie der Sportspiele am Beispiel Tennis*. Frankfurt/M: Deutsch.
- Liu, T., Hohmann, A., Castro, C. & Sörgel, V. (2014 angenommen, zur Veröffentlichung vorgesehen in 2015, s. Bescheinigung). Systematische Spielanalyse im Frauenfußball – Finale des Algarve Cups 2012 Deutschland – Japan. *Zeitschrift für Angewandte Trainingswissenschaft*
- Liu, T. & Hohmann, A. (2013). Applying Data Mining to Analyze the Different Styles of Offense between Manchester United and FC Barcelona in the European Champions League Final. *International Journal of Sports Science and Engineering*, 7(02), 067-078.
- Liu, T. & Hohmann, A. (2013). Applying the Markov Chain theory to Analyze the Attacking Actions between FC Barcelona and Manchester United in the European Champions League final. *International Journal of Sports Science and Engineering*, 7(02), 079-086.
- Loy, R. (1995). Systematische Spielbeobachtung im Fußball. *Leistungssport*, 25(3), 15-20.
- Pfeiffer, M. (2010). *Zusammenfassung der schriftlichen Habilitationsleistung (Unv. Manuskript)*. Bayreuth: Kulturwissenschaftliche Fakultät der Uni-Bayreuth.
- Tamtam, N. (2012). *Data Warehousing und analytische Datenbanken (Unv. Seminararbeit)*. Jena: Universität Jena.
(http://www.informatik.uni-jena.de/dbis/lehre/ws2011/dbsem/Ausarbeitung/A_03_Data_Mining_Naim_Tamtam.pdf)
- Yu, L. J., Zhang, H., & Hu, J. J. (2008). Computer diagnostics for the analysis of table tennis matches. *International Journal of Sports Science and Engineering*, 2 (3), 144-153.
- Zhang, H. (2003). *Leistungsdiagnostik im Tischtennis*. Hamburg: Kovač.

3 Verzeichnis wissenschaftlicher Veröffentlichungen

Zeitschriften (peer-review)

Liu, T., Hohmann, A., Castro, C. & Sörgel, V. (2014 angenommen, zur Veröffentlichung in 2015 vorgesehen). Systematische Spielanalyse im Frauenfußball – Finale des Algarve Cups 2012 Deutschland – Japan. *Zeitschrift für Angewandte Trainingswissenschaft* (s. Bescheinigung).

Liu, T. & Hohmann, A (2013). Applying Data Mining to Analyze the Different Styles of Offense between Manchester United and FC Barcelona in the European Champions League Final. *International Journal of Sports Science and Engineering*, 7(02), 067-078.

Liu, T. & Hohmann, A (2013). Applying the Markov Chain theory to Analyze the Attacking Actions between FC Barcelona and Manchester United in the European Champions League final. *International Journal of Sports Science and Engineering*, 7(02), 079-086.

Abstracts

Liu, T., Hohmann, A., Castro, C. & Sörgel, V. (2013). *Data Mining-gestützte Spielanalyse der deutschen Frauen Nationamannschaft im Finale des Algarve-Cup 2012*. In F. Mess, M. Gruber & A. Woll (Hrsg.), *Sportwissenschaft grenzenlos?! (Schriften der Deutschen Vereinigung für Sportwissenschaft, Band 230, S.135)*. Hamburg: Czwalina.

4 Wissenschaftliche Kongresse

In den letzten drei Jahren habe ich an den folgenden nationalen und internationalen Kongressen teilgenommen und meine Forschungsergebnisse in Form von Vorträgen vorgestellt:

Nationale Kongresse

- 2011: Deutsche Vereinigung für Sportwissenschaft (dvs). Jahrestagung der dvs-Kommission Fußball. Hannover, 24.-26. November 2011
- 2012: Deutsche Vereinigung für Sportwissenschaft (dvs). 22. dvs-Nachwuchsworkshop: Natur- und verhaltenswissenschaftliche Teildisziplinen. Bayreuth, 16.-18. August 2012
- 2013: Deutsche Vereinigung für Sportwissenschaft (dvs). 21. Sportwissenschaftlicher Hochschultag der Deutschen Vereinigung für Sportwissenschaft (dvs). Konstanz, 25. - 27. September 2013

Internationale Kongresse

- 2011: China Sport Science Society (CSSS). 9th China National Congress of Sport Science, Shanghai, China, 5.-7. Dezember 2011.

5 Lehrveranstaltungen

5.1 Beteiligung als Lehrkraft (Projektbetreuer)

- Vorlesung SS 2013:
Trainings- und Bewegungswissenschaft IV (Übung zur VL), Master, (Modul C-3-1), 1 SWS, Hohmann
- Seminar WS 2011/12
Hauptseminar Trainings- und Bewegungswissenschaft (Leistungs- und Wettkampfdiagnostik), (Diplom, Berufsfeldorientierung Leistungssport Bachelor D-10-2, Lehramt S-FW-12-1-2; Hohmann/Singh/Voigt/Liu)
- Seminar SS 2012
Leistungs- und Wettkampfdiagnostik (Berufsfeldorientierung Leistungssport D-10-2, S-FW-12-1-2-, 2 SWS, Hohmann)
- Seminar WS 2012/13:
Hauptseminar Training – Leistung – Wettkampf (Spöko Master C-3-3; Fehr/Voigt)
- Seminar SS 2013:
Leistungs- und Wettkampfdiagnostik (Berufsfeldorientierung Leistungssport D-10-3, B.Sc., S-FW-12-1 LA, 2 SWS, Hohmann)

6 Betreute Abschlussarbeiten und Seminararbeiten

Bachelorarbeiten

1. Systematische Spielbeobachtung im Sportspiel Fußball am Beispiel der deutschen Nationalmannschaft bei der EM 2012 (S. Scholz und M. Kraus)

Seminararbeiten

1. Systematische Spielbeobachtung im Fußball am Beispiel FC Barcelona (C. Dierl und M. Schneider)
2. Diagnose der Wettspielleistung im Fußball - Algarve-Cup 2012 der Frauen (C. Castro und V. Sörgel)
3. Systematische Spielbeobachtung –Traditionell deskriptive Spielbeobachtung im Fußball - ein Vergleich zwischen Champions League Finale 2013 und DFB-Pokal Finale 2012 (N. Kempf, B. Freund und P. Geissler)
4. Systematische Spielbeobachtung im Sportspiel Fußball-DFB Pokalfinale 2012 und Champions League Finale 2013 (G. Knuff)
5. Kommerzielle Spielbeobachtung (A. Schneider)

7 Erklärung

Hiermit versichere ich an Eides Statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt und ist auch noch nicht veröffentlicht.

Bayreuth, den 29.01.2014

(Tianbiao Liu)

Tianbiao Liu

Geburtsdatum: 13. Dezember 1983
Geburtsort: Qinhuangdao, VR China
Familienstand: Ledig
Staatsangehörigkeit: Chinesisch



Bildung und Qualifikation

Seit 04.2011	Promotionsverfahren (kumulativ) an der Universität Bayreuth zum Thema „Sportspielforschung“; Abschluss des Verfahrens im Sommersemester 2014
09.2010 – 01.2011	Studium der deutschen Sprache im Deutschkolleg, Tongji Universität, Shanghai, China
09.2006 – 07.2009	Studium mit Abschluss Master (M.A.) Sportpädagogik und Trainingswissenschaft an der Beijing Sport University
09.2002 – 07.2006	Studium mit Abschluss Bachelor (B.Eng) für Umweltingenieurwesen an der Hebei University of Technology (Technische Universität Hebei)
09.1999 – 07.2002	Nr. 1 Mittelschule, Qinhuangdao

Beruflicher Werdegang

08.2009 – 05.2010	Kent State University China Office, Beijing, China Projektleiter und IT-Leiter, Bereich Marketing
-------------------	--

Praktikum

06.2008 – 07.2009	Praktikum in der Ligaabteilung, des Chinesischen Fußball-Bundes (CFA), Beijing, China
07.2008 – 09.2008	Ehrenamt für Olympische und Paralympische Spiele, Beijing, China

Sportliche Erfahrungen

Seit 08.2012	Fußballspieler des Post SV Bayreuth in der Kreisklasseliga und B – Klasseliga, Bayreuth Landkreis
10.2010	Teilnahme am in „Tianyi Cup“ Fußballturnier, Shanghai.

Auszeichnung

08.2010	DAAD-Doktorandenstipendium
09.2008	„Outstanding Olympic Volunteer“, BOCOG, Beijing, China

Bayreuth, den 29.01.2014

Tianbiao Liu

Applying Data Mining to Analyze the Different Styles of Offense between Manchester United and FC Barcelona in the European Champions League Final

Tianbiao Liu ⁺, Andreas Hohmann

Institute of Sports Science, University of Bayreuth, 95447 Bayreuth, Germany

(Received June 4, 2012, accepted September 9, 2012)

Abstract. Soccer games are always analyzed by means of traditional statistical methods. With the development of modern IT technology, new methods were introduced to analyze soccer games. In this study of the European Champions league final between FC Barcelona and Manchester United, detailed game actions were recorded and data mining methods were applied. Microsoft SQL Server Data mining add-ins served to calculate association rules and dependency networks. The data mining results led to typical playing patterns of both teams.

Keywords. Soccer, European Champions league final, Data mining, Association rules, Dependency networks

1. Introduction

In soccer games, offense plays a key role in regard to scoring goals and winning the game. To analyze the different attacking styles of soccer teams, researchers have developed various methods that are based on traditional statistical procedures [1],[2]. In recent years, researchers introduced advanced mathematical models for soccer game observation and research strategies. To evaluate team characteristics, like the optimal timing of substitutions and tactical decisions, Hirotsu and Mike Wright [3]; [4] made use of Markov's chain theory. In addition, during the 2006 FIFA World Cup®, team tactical features were researched by Pfeiffer, Hohmann and Buehrer [5]. At present, increasingly large amounts of data in sports are collected, which creates a need for innovation in research methods. Since the use of database became an integral part in sports research, data mining methods (e.g. association rule) were first applied in table tennis [6]. In football games, association rules models have been introduced to analyze football techniques [7] and tactics [8], but most research projects focused on the development of data mining algorithm instead of its application.

In this case study, data mining methods (model of association rules) were applied to analyze the tactical behavior during all phases of ball possession for both teams (Manchester United and FC Barcelona) in the European Champions League final 2011.

2. Methods

2.1. Data collection

The match took place on May 29th, 2011, and was recorded by Video so that all game elements during the ball possession phases of the two opposing teams could be edited afterwards and used to a) analyze the game and b) compare the tactical behavior during offense.

2.2 Division of soccer field areas

As shown in Figure 1, the whole playing field is divided into 30 zones. In the first half, Manchester United attacked from right to left so their attacking area included zones #1-10. The midfield area consisted of

⁺ Tianbiao Liu. Tel.: +49-921-55 3479; fax: +49-921-55 5806
E-mail address: ltbvane@yahoo.com.cn.

zones #11-20, and the backfield area of zones #21-30. Each of the three areas covered the same amount of space, 35 m out of a total of 105 m. In contrast, FC Barcelona attacked from left to right so their attacking area referred to zones #21-30, the midfield area consisted of zones #11-20, and the backfield area represented zones #1-10. In order to facilitate the recording after the teams changed ends at halftime, the numbering of the zones of the playing field remained unchanged.

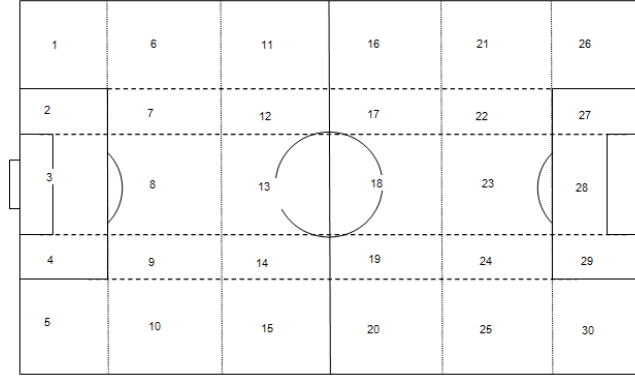


Fig. 1: Spatial division of soccer field areas

In addition to the three horizontal zones mentioned, a vertical stripe was created by zones #2-4, 7-9, 12-14, 17-19, 22-24, and 27-29. Furthermore, on both sides of this longitudinal pathway two side wings were identified by zones #1, 6, 11, 16, 21, 26, and # 5, 10, 15, 20, 25, and 30, respectively.

2.3 Data mining

2.3.1 Datamining theory

Data mining refers to extracting or “mining” knowledge from large amounts of data [9]. It is the kind of technology which extracts the non-ordinary, implicit, unknown, potentially useful information from the large-scale data [1],[11].

Data mining involves six common classes of tasks [12]: (1) anomaly detection, (2) association rule learning, (3) clustering, (4) classification, (5) regression, and (6) summarization. The association rule procedure has been used in sport events analysis to detect frequent playing patterns [13], since teams or players tend to exhibit relatively stable playing patterns. For example, teams may have players who focus on organizing the attacks, and forwards who often create one on one situation by heading into certain areas of the attacking area.

2.3.2 Association rule and Apriori algorithm

The problem of association rule mining is defined as: Let $I = \{i_1, i_2, \dots, i_m\}$ be a set of binary attributes, called items. Let D be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. Associated with each transaction is a unique identifier, called its TID. We say that a transaction T contains X , a set of some items in I , if $X \subseteq T$. An association rule is an implication of the form $X \Rightarrow Y$, where $X \subseteq I, Y \subseteq I$, and $X \cap Y = \emptyset$. The rule $X \Rightarrow Y$ holds in the transaction set D with confidence c if $c\%$ of transactions in D contain X also contain Y . The rule $X \Rightarrow Y$ has support s in the transaction D if $s\%$ of transactions in D contains $X \cup Y$ [14].

Many efficient and scalable algorithms have been developed for frequent item set mining and the Apriori algorithm is a seminal and basic algorithm for Boolean association rules [xv]. Other improved algorithms all come from the basic Apriori. The Apriori employs an iterative approach known as a level-wise search, where k -item sets are used to explore $(k+1)$ -item sets. First, the set of frequent 1-item sets is found by scanning the database to accumulate the count for each item, and collecting those items that satisfy minimum support. The resulting set is denoted L_1 . Next L_1 is used to find L_2 , the set of frequent 2-item sets, which is used to find L_3 , and so on, until no more frequent k - item sets can be found. The finding of each L_k requires one full scan

of the database.¹

To improve the efficiency of the level-wise generation of frequent item sets, an important property called the a priori property is used to reduce the search space: *All nonempty subset of a frequent item set must also be frequent.*

2.3.3 Data structure

During ball possession phases the players dribble and keep passing the ball to each other, which are important actions of ball movement into the offensive area in front of the opponent's goal. Therefore, the main interest of this study was to find out which players are important for each team's playing style, and which players serve as critical passing connections in the tactical stream of actions of the two teams.

In our research, every piece of detailed information of the game flow on the field was recorded. A team's control began when the team took possession of the ball (kick-off, free kick, throw-in, corner kick, goal kick, goalkeeper throw of the ball after the opponent's attack, possession of the ball from a tackle or one on one situation), and ended with a shot on the goal, foul, or loss of control of the ball. The data structure was as follows:

{Offensive sequence; Event in the sequence; Player number; Zone;}

The offensive sequence is the number of an offensive chain. An event in the sequence is each action number in the chain. The player number is the player who controls the ball at the moment. The zone is where the ball appears at the moment. As this structure is built, data mining can be processed to discover interesting features.

2.3.4 Application of the data mining tool



Fig. 2: Microsoft Data Mining Add-ins for Excel 2007

Table 1: part of game database using vertical format

Offensive sequence -SID	Event in the sequence - EID	Player NO. – Item set	Zone – Item set
1	1	22	6
1	2	10	11
1	3	16	12
1	4	2	14
1	5	2	20
2	1	1	3
2	2	14	5
2	3	14	10
2	4	1	3
2	5	14	10
3	1	22	11
3	2	16	11
3	3	22	11
4	1	22	6
4	2	16	6

¹ For detailed information about Apriori algorithm see Data mining: Concepts and Techniques (Han& Kamber, 2006).

A controlling chain flow (attacking flow) was developed to analyze the two team's playing patterns. In this case Microsoft Data Mining Add-ins for Excel 2007 was applied (see Fig. 2). To match the mining tool vertical data format (see Tab. 1) was applied.

In Microsoft association rule algorithm, there are three parameters reflecting the rules:

Support, Probability, and Importance

Support, which is sometimes referred to as frequency, means the number of cases that contain the targeted item or combination of items. Only items that have at least the specified amount of support can be included in the model. In contrast, the threshold for rules is expressed not as a count or percentage, but as a **probability**, sometimes referred to as confidence (see 2.3.2). Besides support and confidence (here in MS tool called probability) for each item set, the algorithm then creates scores that represent support and confidence. These scores can be used to rank and derive interesting rules from the item sets, which indicates its **importance** [15].

3. Results

3.1. Objectivity of the game observation models and model validity

In this study, all match events included in the game analysis were examined in regard to inter-rating consistency of two observers (inter-observer agreement) that was quantified by Cohen's Kappa. Manchester United first half record was selected for the examination. The Cohen's kappa values (κ) of the models were found to be: $\kappa = 0.766$ for "player's number" and $\kappa = 0.625$ for "zone". These values indicate that κ is sufficient and worthwhile to use [16], [17]. However, since the κ -values are not perfect, the difficulties of collecting match data from video tapes are underlined.

3.2. Applying Association rule to analyze both teams' control and offensive actions

After establishing the association rules model, the match data of Manchester United were mined for the first half of the game. In the data mining process, the threshold value of minimum *support* was given 5%; the threshold of minimum *probability* (confidence) was fixed at 50 %.

3.2.1 Player passing model for Manchester United (First half)

As shown in Table 2, high probabilities appeared mostly with the two center backs (#5, Ferdinand, and #15, Vidic) meaning that when one center back held the ball and then passed it to his teammate, it appeared that the ball would come back to another center back within only a few steps. The high value of importance indicates that these rules are of good quality.

Table 2: Association rule: Player number, which suggests player connections

Probability	Importance	Rule
83 %	0.63	5 = Existing, 1 = Existing -> 15 = Existing
83 %	0.63	15 = Existing, 1 = Existing -> 5 = Existing
71 %	0.30	5 = Existing, 15 = Existing -> 1 = Existing
64 %	0.63	15 = Existing -> 5 = Existing
64 %	0.63	5 = Existing -> 15 = Existing
64 %	0.29	10 = Existing -> 11 = Existing
60 %	0.24	13 = Existing -> 11 = Existing
56 %	0.46	25 = Existing -> 15 = Existing
56 %	0.46	25 = Existing -> 5 = Existing
56 %	0.19	25 = Existing -> 1 = Existing
56 %	0.19	25 = Existing -> 11 = Existing
55 %	0.20	5 = Existing -> 1 = Existing
55 %	0.20	15 = Existing -> 1 = Existing
50 %	0.41	13 = Existing -> 3 = Existing
50 %	0.30	20 = Existing -> 16 = Existing
50 %	0.14	20 = Existing -> 11 = Existing

Figure 3 shows clearly and directly Manchester United's control of the ball in the first half of the match. From the chart one can see that the midfield players of Manchester United tend to have more vertical connections with their full backs instead of passing patterns between the midfield players themselves. That means that when one Manchester United midfield player possessed the ball and passed out, after a few steps the ball tended to go back to Manchester United defenders instead of being continuously controlled by the midfield players. One reason for that could be that the Manchester United midfield was limited by the tactical play of FC Barcelona disrupting the flow of passes which left the Manchester United midfield players with mainly one other option: vertical passes.

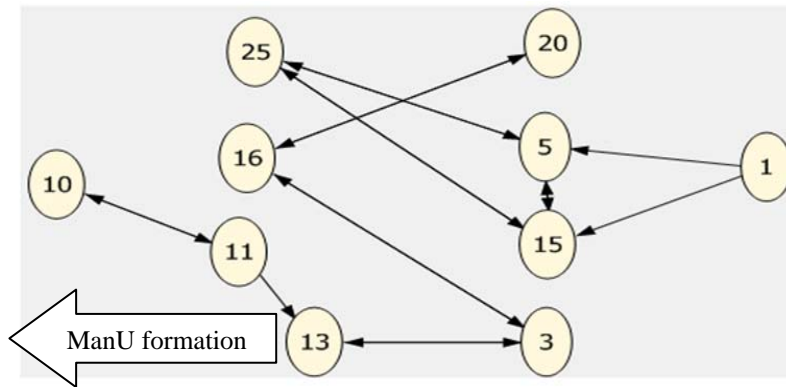


Fig. 3: Dependency network: Player passing trend of Manchester United (first half)

3.2.2 Ball movement model from area to area for Manchester United (First half)

Table 3: Association Rules: Connection between zones for Manchester United (First half)

Probability	Importance	Rule
100 %	0.77	9 = Existing -> 10 = Existing
75 %	0.67	8 = Existing -> 13 = Existing
50 %	0.35	22 = Existing -> 28 = Existing
50 %	0.52	14 = Existing -> 13 = Existing
50 %	0.35	17 = Existing -> 18 = Existing
50 %	0.60	16 = Existing -> 11 = Existing
50 %	0.60	17 = Existing -> 11 = Existing
50 %	0.38	19 = Existing -> 28 = Existing
50 %	0.38	19 = Existing -> 18 = Existing
50 %	0.52	17 = Existing -> 13 = Existing

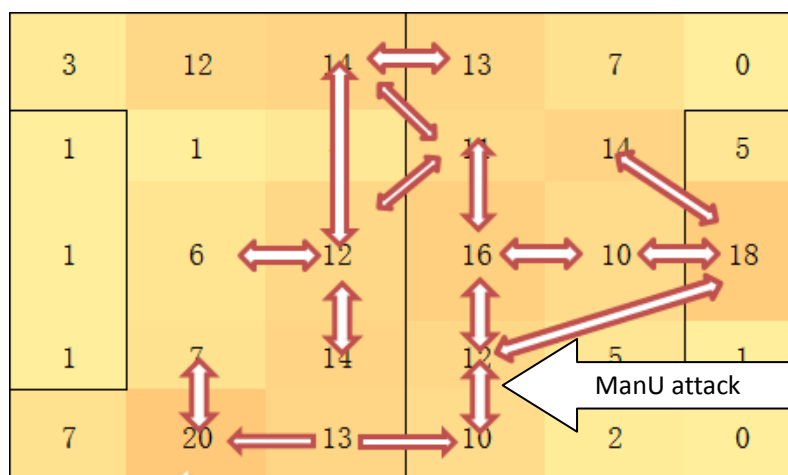


Fig. 4: Dependency Network: Zones transition for ManU (First half)

Table 3 of the first half shows that when Manchester United players controlled the ball in area #9, within the next few steps, the ball would most likely move to area #10 which turned out to be of high quality. (Importance = 0.77).

The corresponding dependency network (see Fig. 4) demonstrates more clearly the tactical differences in the Manchester United's preferred ball controlling areas. The ball could only be passed more transversely in their own half of the field compared to the number of passes into the front zones of the attacking area. This lack of vertical play was even more evident, when the ball possession took place in the front zones not closer than 35 m to Barcelona's goal.

3.2.3 Player passing model for Manchester United (Second half)

Table 4: Association Rule: Player connections for ManU (Second half)

Probability	Importance	Rule
75 %	0.25	25 = Existing -> 10 = Existing
71 %	0.52	18 = Existing -> 15 = Existing
71 %	0.46	18 = Existing -> 25 = Existing
67 %	0.17	11 = Existing -> 10 = Existing
64 %	0.58	15 = Existing -> 1 = Existing
64 %	0.58	1 = Existing -> 15 = Existing
64 %	0.28	13 = Existing -> 3 = Existing
55 %	0.44	15 = Existing -> 13 = Existing
55 %	0.44	15 = Existing -> 5 = Existing
55 %	0.44	5 = Existing -> 15 = Existing
55 %	0.44	13 = Existing -> 15 = Existing
55 %	0.38	5 = Existing -> 25 = Existing
55 %	0.05	13 = Existing -> 10 = Existing
50 %	0.40	25 = Existing -> 5 = Existing
50 %	0.13	11 = Existing -> 3 = Existing

Table 4 shows some part of association rules of players' tactical connections within the team of Manchester United. The passes from the players #25 to #10, #18 to #15, and #18 to #25 exhibit the highest probabilities (confidence). Figure 5 presents more clearly the players connections during the second half of the match. According to this chart, Rooney played the role of the target man in Manchester United's front zone attacks.

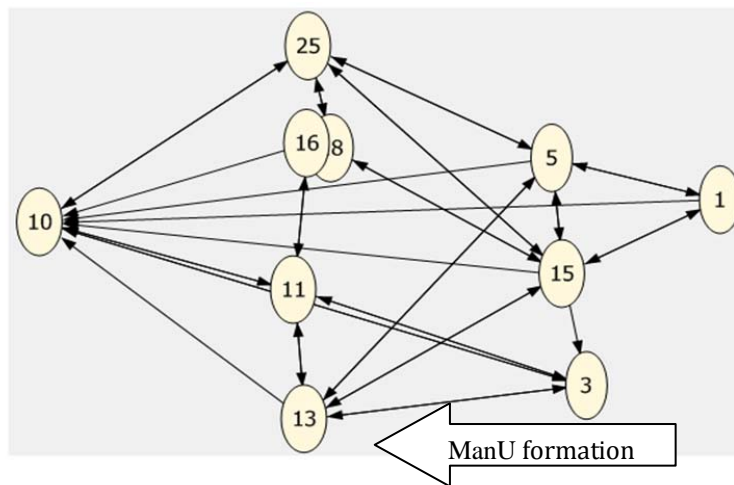


Fig. 5: Dependency Network: player passing trend of ManU (Second half)

3.2.4 Ball movement model from area to area for Manchester United (Second half)

Table 5: Association Rules: Connection between zones for Manchester United second half

Probability	Importance	Rule
100 %	0.72	11 = Existing -> 16 = Existing
75 %	0.38	17 = Existing -> 18 = Existing
71 %	0.89	8 = Existing -> 9 = Existing
71 %	0.89	9 = Existing -> 8 = Existing
63 %	0.53	17 = Existing -> 16 = Existing
60 %	0.27	16 = Existing -> 18 = Existing
60 %	0.27	15 = Existing -> 18 = Existing
50 %	0.60	16 = Existing -> 17 = Existing
50 %	1.20	16 = Existing -> 11 = Existing

Table 5 shows the data mining results of the passing between the playing field zones for Manchester United in the second half. The highest probability (confidence) occurs for passes from area #11-#16 with decent importance (0.72), while the highest importance is registered for passes from area #16-#11 with a probability of 0.5. Figure 6 presents the dependency network of the ball possession areas in the second half of the match for the ManU team. From the hardly existing connections (as illustrated in the chart), it is easy to see that Manchester United was severely limited in its attacks during the second half.

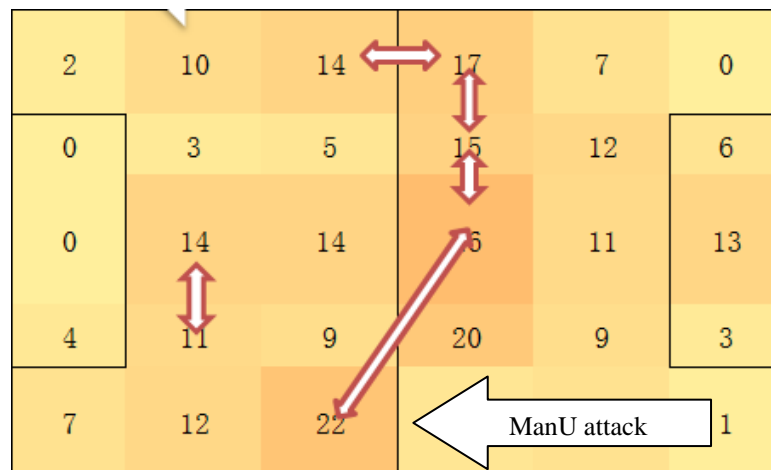


Fig. 6: Dependency network: zones transition for ManU second half

3.2.5 Player passing model for FC Barcelona (First half)

Table 6: Association Rule: Player connections for FC Barcelona (First half)

Probability	Importance	Rule
100 %	0.22	17 = Existing, 16 = Existing -> 6 = Existing
94 %	0.36	2 = Existing, 6 = Existing -> 10 = Existing
94 %	0.29	2 = Existing, 10 = Existing -> 6 = Existing
93 %	0.29	2 = Existing, 16 = Existing -> 10 = Existing
93 %	0.23	2 = Existing, 8 = Existing -> 6 = Existing
93 %	0.29	2 = Existing, 8 = Existing -> 10 = Existing
93 %	0.22	17 = Existing, 8 = Existing -> 6 = Existing
93 %	0.30	17 = Existing, 6 = Existing -> 8 = Existing
92 %	0.25	7 = Existing -> 10 = Existing
92 %	0.19	3 = Existing, 16 = Existing -> 6 = Existing
91 %	0.24	17 = Existing, 10 = Existing -> 8 = Existing

Table 6 shows parts of the association rules of the FC Barcelona players' passing trend. When #17 (Pedro) had the ball, within a few passes, #16 Busquets most probably got the ball, and then midfielder #6 (Xavi) had the highest possibility of receiving the ball.

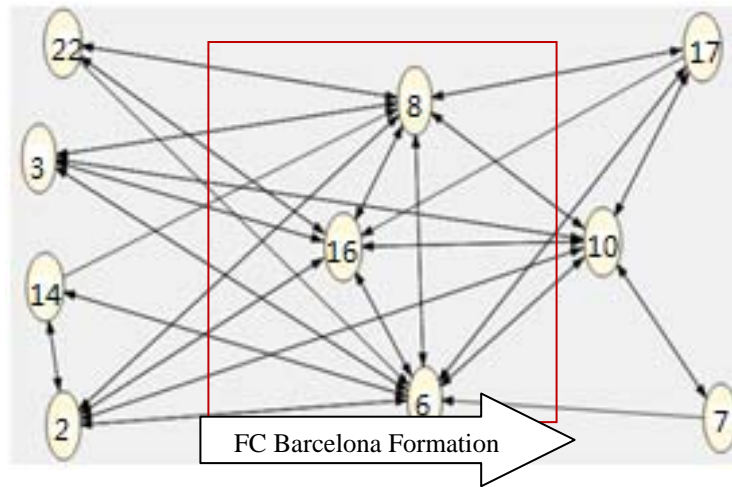


Fig. 7: Dependency network: player passing trend of FCB (First half)

Figure 7 shows FC Barcelona's connection trend. The FC Barcelona midfielders played important roles in the team's control and attack. Both offensive players (#17 and #7) on the wing sides had connections mainly with midfield players, and the fullback players were primarily connected to midfielders, too, but in no way directly to the forwards (#7 and #17). The midfielders #6 (Xavi), #8 (Iniesta) and #10 (Messi) organized the attack in a way that is typical for FCB's playing pattern.

3.2.6 Ball moving model from area to area for FC Barcelona (First half)

Table 7: Association Rules: Connection between zones for FCB (First half)

Probability	Importance	Rule
72 %	0.51	13 = Existing -> 18 = Existing
72 %	0.51	18 = Existing -> 13 = Existing
67 %	0.35	23 = Existing -> 18 = Existing
56 %	0.41	18 = Existing -> 23 = Existing

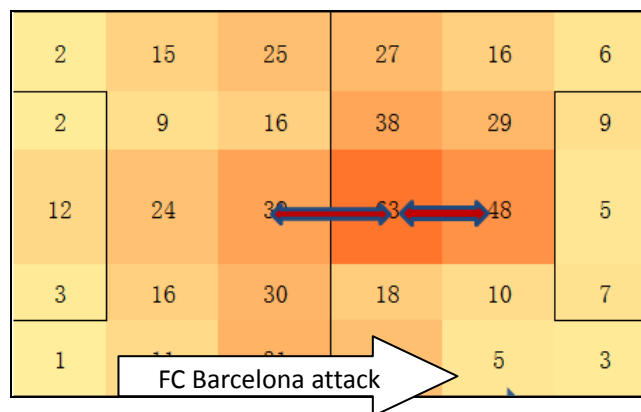


Fig. 8: Dependency Network: Zones transition for FCB first half

Table 7 shows the data mining results of the zones connections in the FC Barcelona first half match play, where the areas #13 and #18; #18 and #13; #23 and #18; #18 and #23 are characterized by high probabilities (confidence) of ball possessions. Hence the association rules for passes from area #13 to #18 and from #18 to #13 show a relative high quality. Figure 8 presents the dependency network of the different playing zones for the FC Barcelona team in the first half of the match. The chart demonstrates the FC Barcelona's controlling and passing trends that were predominantly focused on the central positions.

3.2.7 Player passing model for FC Barcelona (second half)

Table 8: Association Rule: Player connections for FCB (Second half)

Probability	Importance	Rule
100 %	0.32	22 = Existing, 6 = Existing -> 10 = Existing
100 %	0.29	8 = Existing, 2 = Existing -> 6 = Existing
94 %	0.32	8 = Existing, 10 = Existing -> 6 = Existing
91 %	0.23	22 = Existing, 10 = Existing -> 6 = Existing
91 %	0.28	8 = Existing, 2 = Existing -> 10 = Existing
89 %	0.33	8 = Existing -> 6 = Existing
88 %	0.35	8 = Existing, 6 = Existing -> 10 = Existing
88 %	0.28	2 = Existing, 10 = Existing -> 6 = Existing
85 %	0.26	22 = Existing -> 10 = Existing
85 %	0.43	10 = Existing -> 6 = Existing
84 %	0.34	8 = Existing -> 10 = Existing
79 %	0.28	2 = Existing -> 6 = Existing
79 %	0.28	2 = Existing, 6 = Existing -> 10 = Existing
79 %	0.51	6 = Existing -> 10 = Existing
77 %	0.14	22 = Existing -> 6 = Existing

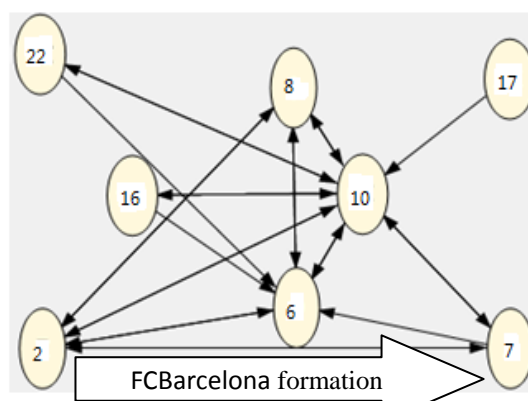


Fig. 9: Dependency network: player passing trend of FCB (Second half)

Table 8 and Figure 9 display the data mining results of the player passing model for FC Barcelona in the second half of the match. The triplets of passes from #22 to #6 to #10, and from #8 to #2 to #6 represent the main passing trends. The dependency network clarifies the tactical combinations among the players. Besides the typical playing style along the vertical midfield axis, combinations on the right wing occurred more often.

3.2.8 Ball movement model from area to area for FC Barcelona (Second half)

Table 9 and Figure 10 show the data mining results for FC Barcelona in the second half of the match. The passing trends from zone #15 to #14, #19 to #18, and from #20 to #18 occur with highest probability, indicating that these three rules are of relative high importance. In the dependency network (see Fig. 10) it can be seen that the tactical connections between FC Barcelona's midfield zones were still strong, and the right side appeared in more cases compared to the first half of the match. This finding is supported by the analysis in chapter 3.2.7, which points out that player #2 was very active on the right side in the second half of the match.

Table 9: Association Rules: Connection between zones for FCB (Second half)

Probability	Importance	Rule
86 %	0.61	15 = Existing -> 14 = Existing
85 %	0.56	19 = Existing -> 18 = Existing
83 %	0.51	20 = Existing -> 18 = Existing
77 %	0.47	19 = Existing -> 14 = Existing
72 %	0.56	14 = Existing -> 18 = Existing
72 %	0.56	18 = Existing -> 14 = Existing
71 %	0.38	12 = Existing -> 13 = Existing
71 %	0.43	15 = Existing -> 18 = Existing
67 %	0.83	14 = Existing -> 15 = Existing
67 %	0.40	18 = Existing -> 13 = Existing
63 %	0.42	13 = Existing -> 18 = Existing
61 %	0.79	18 = Existing -> 19 = Existing
56 %	0.75	18 = Existing -> 20 = Existing
56 %	0.63	14 = Existing -> 19 = Existing
56 %	0.23	14 = Existing -> 13 = Existing
56 %	0.53	18 = Existing -> 15 = Existing
53 %	0.50	13 = Existing -> 12 = Existing
53 %	0.24	13 = Existing -> 14 = Existing

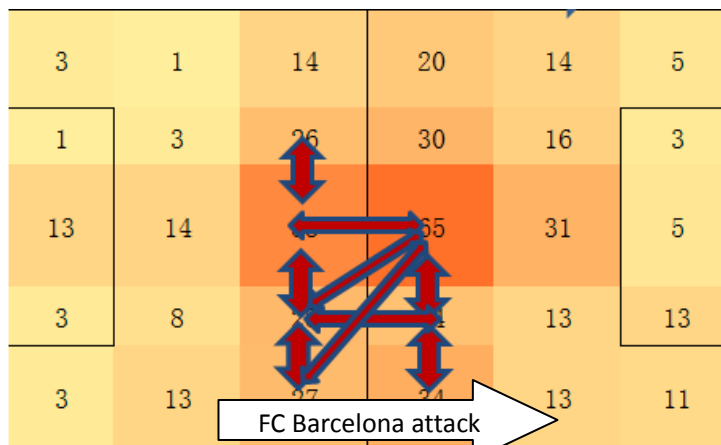


Figure 10 Dependency Network: Zones transition for FCB (Second half)

4. Discussion and conclusion

In this study, data mining theory was applied to analyze the attacking actions of the two opposing teams FC Barcelona and Manchester United in the European Champions League final 2011. By using an association rules data mining tool, both teams' basic playing patterns during their ball possession phases were clearly shown through dependency networks.

Obviously, FC Barcelona relied much more on its midfield players from the beginning of the match until the very end. Especially in the first half Manchester United's midfield was thoroughly kept in check. Although Manchester United changed its attacking strategy after the break and focused more on its forward Rooney, the English midfield vows continued as the team of FC Barcelona stayed in firm control of its midfield until the end of the match.

5. References

- [1] Kuhn, Werner and Schmidt, Werner. *Analyse und Beobachtung in Training und Wettkampf: Beiträe und Analysen zum Fussballsport IV*. Sankt Augustin: Academia Verlag, 1991.
- [2] Jens. Bangsbo, Tomas. Reilly, and Charles. Hughes. *Science and football*. Spon Press, 1997.

- [3] N. Hirotsu, and M. Wright. An evaluation of characteristics of teams in association football by using a Markov process model. *The Statistician*. 2003, **52**: Part 4, pp. 591–602
- [4] N. Hirotsu, and M. Wright. Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions. *Journal of the Operational Research Society*. 2002, **53**: 88-96.
- [5] M. Pfeiffer, A. Hohmann, and M. Buehrer. Computersimulation zur Bestimmung der Leistungsrelevanz taktischer Verhaltensweisen bei der FIFA WM 2006.5. Dvs-sportspiel-Symposium Universitaet Flensburg, 2006.
- [6] L. Yu, H. Zhang, and J. Hu. Computer diagnostics for the analysis of table tennis matches. *International Journal of Sports Science and Engineering*. 2008, **2**(03): 144-153.
- [7] B. Wang, Z. Yin, and L. Wang. Research of Association Rules in Analyzing Technique of Football Match. 2nd *International Conference on Power Electronics and Intelligent Transportation System*. 2009.
- [8] C. Pan. Appliance of Apriori Algorithm on technical-tactics analysis of football. *Computer Knowledge and Technology*. 2010, **31**(6): 8835-8837.
- [9] J. Han, and M. Kamber. *Data Mining: Concepts and Techniques* (Second Edition), Elsevier Inc, 2006.
- [10] J. Brachman, and T. Anand. The Process of Knowledge Discovery in Databases. *A Human centered Approach*. 1996, pp. 56-60.
- [11] P. Smyth. From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence*. 1991, pp. 56-60.
- [12] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From Data Mining to Knowledge Discovery in Databases. Retrieved 2008-12-17.
- [13] I. Bhandari, E. Colet, and J. Parker. Advanced Scout: Data Mining and Knowledge Discovery in NBA Data. *Data Mining and Knowledge Discovery*. 1997, **1**: 121–125,
- [14] Agrawal and Srikant. *Fast Algorithms for Mining Association Rules*. 1994.
- [15] J. Han, and M. Kamber. *Data Mining: Concepts and Techniques* (Second Edition). Elsevier Inc, 2006.
- [16] MSDN, Microsoft Association Algorithm Technical Reference. <http://msdn.microsoft.com/en-us/library/cc280428.aspx>
- [17] W. Greve, and D. Wentura. *Wissenschaftliche Beobachtung: Eine Einführung*, PVU/Beltz, Weinheim 1997.
- [18] J. Landis, and G. Koch. The measurement of observer agreement for categorical data. In: *Biometrics*. 1977, **33**: 159–174.

Applying the Markov Chain theory to Analyze the Attacking Actions between FC Barcelona and Manchester United in the European Champions League final

Tianbiao Liu⁺, Andreas Hohmann

Institute of Sports Science, University of Bayreuth, 95447 Bayreuth, Germany

(Received June 27, 2012, accepted September 9, 2012)

Abstract. Several researchers have applied Markov chain methods in sport game analysis. This study analyzed the European Champions League final between Manchester United and FC Barcelona based on a first order Markov chain model. The results describe important passing strategies and tactical connections between certain players from each team.

Keywords. Soccer, European Champions League final, Markov chain, Stochastic process

1. Introduction

For a long time, researchers have shown interest in studying offensive actions in soccer games, and as a result, lots of methods and tools in soccer game analysis were developed. The majority of them are based on traditional statistical procedures [1], [2]. Later researchers incorporated mathematical methods into sports game research. The Markov Chain theory has already been applied to analyze Table tennis matches [3]. Concerning a soccer game, making use of the Markov Chain theory, the optimal timing of substitution and tactical decisions [4], as well as evaluation of team characteristics were determined by Hirotsu and Mike Wright [5]. In addition, during the 2006 FIFA World Cup team tactical features were researched by Pfeiffer, Hohmann and Buehrer [6].

In this case study, the Markov chain theory was applied to analyze offensive actions for both teams (Manchester United and FC Barcelona) in the 2011 European Champions League Final.

2. Methods

2.1. Data collection

The game was held on May 29th, 2011 and was videotaped so that all elements could be marked afterwards. Every detail of the game was observed and used for analysis to be able to draw comparisons between the two great teams.

2.2. Division of soccer field areas

As shown in Figure 1, the whole playing field was divided into 30 zones. In the first half, Manchester United attacked from right to left in the first half so their attacking area included zones #1-10. The midfield area consisted of zones #11-20, and the backfield area of zones #21-30. Each area covered the same amount of space, 35m out of 105m. In contrast, FC Barcelona attacked from left to right so their attacking area referred to zones #21-30, their midfield area to zones #11-20, and their backfield area to zones #1-10. In order to facilitate the recording after the teams changed ends at halftime, the numbering of the zones remained unchanged.

In addition, middle zones were defined by # 2-4, 7-9, 12-14, 17-19, 22-24, and 27-29. Side zones were those besides middle zones.

⁺ Tianbiao Liu. Tel.: +49-921-55 3479; fax: +49-921-55 5806
E-mail address: ltbvane@yahoo.com.cn.

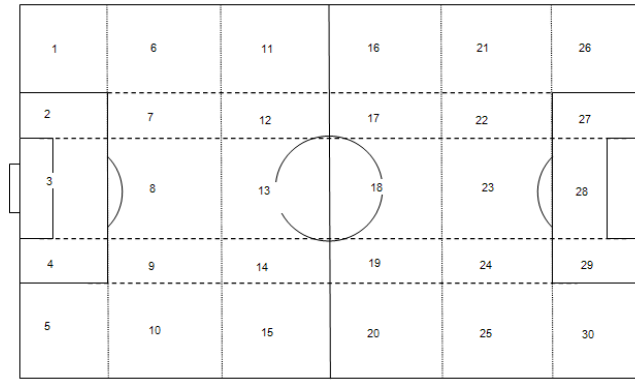


Fig. 1: Division of soccer field areas

2.3. Markov Chain

2.3.1 Markov chain theory

Mathematically, a Markov chain is a discrete random process with the Markov property. A discrete random process means a system which can be in various states, and which changes randomly in discrete steps.

A Markov chain is a sequence of random variables X_1, X_2, X_3, \dots with the Markov property, namely that, given the present state, the future and past states are independent. Formally,

$$\Pr(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \Pr(X_{n+1} = x | X_n = x_n)$$

The possible values of X_i form a countable set S called the state space of the chain.

Markov chains are often described by a directed graph, where the edges are labeled by the probabilities of going from one state to the other states.

The changes of the states of the system are called transitions, and the probabilities associated with various state-changes are called transition probabilities.

2.3.2 Observation model

The main purpose of a soccer team is to score goals, specifically, score more goals than its opponent. In order to be successful, players organize attacks as soon as they take control of the ball from their opponent. Next, they keep on dribbling and passing the ball to each other until they lose control of the ball or other scenarios as selected by the researchers. In this case, the attacking/backfield third (35m in front of the goal) was selected. The kind of chain system is applied in this research as an analysis unit, and the observation unit (state) is defined as the player who controls the ball in the chain. The objectivity of the observation model is confirmed by the agreement of two independent observers by using Cohen's Kappa statistic [7].

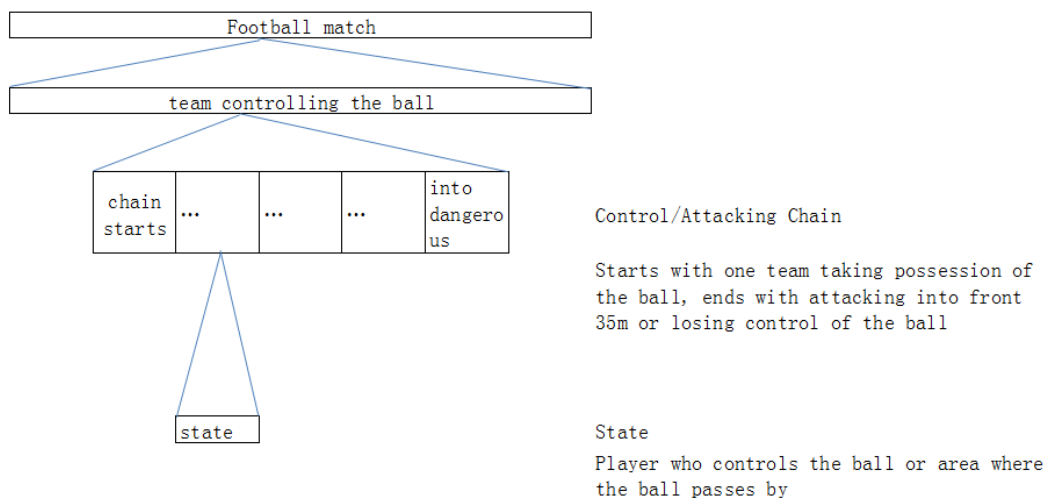


Fig. 2: Data structure of football game

2.3.3 Stochastic model

The transition probabilities between two states describe the soccer match as a process that can be understood as a first order Markov Chain. The following two properties are given: a) the probability for the next state depends solely on the current state (Markov property), it has nothing to do with any earlier state (player); and 2) the transition probability from one state to another is independent of its chronological position in the match process (Chain-property).

The transition probabilities between the states can be transformed into a two-dimensional transition matrix. Each element in the matrix $P_{ij}(n) \geq 0$ and the sum of line is equal to 1 [8]. In the Markov chain, absorbing states are important, since the process ends in these states and a new process starts in these states. In this current research, the state “the front 35m” is defined as the absorbing state. One can calculate the Attacking Probability (AP) by using all the transition probabilities (see Table 1).

Table 1: Example of transition matrix by using observation model “Passes” (ManU in the first half)

	Van Der Sar 1	Patrice Evra 3	Rio Ferdinand 5	Wayne Rooney 10	Ryan Giggs 11	Park Ji-Sung 13	Javier Hernandez 14	Nemanja Vidic 15	Michael Carrick 16	Fabio 20	Antonio Valencia 25	into 35m 35
Van Der Sar 1	5.00	5.00	15.00	5.00	5.00	5.00	0.00	0.00	0.00	5.00	5.00	5.00
Patrice Evra 3	0.00	19.05	0.00	9.52	4.76	28.57	4.76	4.76	14.29	0.00	0.00	4.76
Rio Ferdinand 5	21.05	10.53	10.53	0.00	5.26	5.26	0.00	15.79	15.79	0.00	5.26	5.26
Wayne Rooney 10	0.00	3.45	3.45	17.24	13.79	6.90	6.90	0.00	10.34	6.90	6.90	13.79
Ryan Giggs 11	0.00	2.94	5.88	5.88	23.53	11.76	2.94	2.94	2.94	8.82	2.94	8.82
Park Ji-Sung 13	0.00	16.67	0.00	16.67	4.17	12.50	0.00	4.17	4.17	4.17	0.00	20.83
Javier Hernandez 14	0.00	0.00	0.00	15.39	7.69	7.69	23.08	0.00	23.08	0.00	7.69	15.38
Nemanja Vidic 15	13.04	4.35	26.09	4.35	13.04	4.35	0.00	13.04	8.70	4.35	0.00	0.00
Michael Carrick 16	4.17	0.00	0.00	8.33	12.50	0.00	12.50	4.17	12.50	20.83	0.00	4.17
Fabio 20	0.00	0.00	8.71	13.04	21.74	0.00	0.00	0.00	0.00	13.04	17.39	4.35
Antonio Valencia 25	0.00	0.00	13.33	6.67	6.67	0.00	0.00	13.33	6.67	20.00	6.67	6.67

2.3.4 Calculating the performance relevance

Based on the ball moving matrix and player passes matrix above, it is possible to calculate the AP on the basis of a simulated transition matrix. In order to determine the performance relevance of a tactical behavior pattern, each cell in the initial matrix will be firstly modified by a certain percentage resulted from a function:

$$\delta TP = C + B \times 4 \times TP(1 - TP) \quad (\text{Lames, 1991}).$$

In this function, TP is the transition probability; δTP is the change of element transition probability. The constant values applied in the study are $C = 1$, $B = 5$, which were determined by Lames [9] and tested by Pfeiffer [10].

	Van Der Sar_1	Patrice Evra_3	Rio Ferdinand_5	Wayne Rooney_10	Ryan Giggs_11	Park Ji-Sung_13	Javier Hernandez_14	Nemanja Vidic_15	Michael Carrick_16	Fabio_20	Antonio Valencia_25	into 35m	out of 35m
Van Der Sar_1	5.00	5.00	15.00	5.00	5.00	5.00	0.00	0.00	0.00	5.00	5.00	5.00	45.00
Patrice Evra_3	0.00	19.05	0.00	9.52	4.76	28.57	4.76	4.76	14.29	0.00	0.00	4.76	9.52
Rio Ferdinand_5	21.05	10.53	10.53	0.00	5.26	5.26	0.00	15.79	15.79	0.00	5.26	5.26	5.26
Wayne Rooney_10	0.00	3.45	3.45	17.24	13.79	6.90	6.90	0.00	10.34	6.90	6.90	13.79	10.34
Ryan Giggs_11	0.00	2.94	5.88	5.88	23.53	11.76	2.94	2.94	2.94	8.82	2.94	8.82	20.59
.....													
AP _{into 35m} (empirical matrix) = 30.31%													
AP _{into 35m} (manipulated matrix) = 30.29%													
$\delta AP_{\text{into 35m}} = -0.02\%$													

Fig. 3: Example of simulative calculation of the Attacking relevance (δAP) of model “passes”

In order to keep the line sum in the matrix still equal to 1.00, the other values in the same row must be proportionally compensated; in this case, a compensation function is introduced and applied:

$$\delta TP_{xi} = - \left[\frac{TP_i}{(1 - TP_i)} \right] \times \delta TP_x$$

After this, another AP is calculated from the new value in the cell which helps to define the performance relevance (δAP) of a tactical behavior pattern in terms of the difference between the attacking probability (AP) as calculated by the original transition-matrix and the attacking probability as calculated by the modified transition-matrix (see Figure 3).

The higher the δAP is, the more positively the element (game state transition) affects the game, correspondingly, the lower the δAP is, the more negatively the element (game state transition) affects the game.

Similarly, all other δAP s can be calculated and comparisons can be made which are used to identify which passes (giving or receiving) are more efficient and also effective in the game.

3. Results

3.1. Objectivity of the game observation models and model validity

In this study, all match events were recorded and included into analysis. Their consistency was examined by the inter-rating consistency of two observers (inter observer agreement) quantified in Cohen's Kappa. Manchester United first half record was selected for examination of Cohen's kappa. The Cohen's kappa values (κ) of the model were found to be: $\kappa = 0.766$ for "player's number", $\kappa = 0.625$ for "Zone", indicating their usefulness [11], [12]. But these κ values also reflected the problems to identify all data while collecting them from TV.

3.2. Analysis of offensive actions

3.2.1 Performance relevance of both teams' tactical behaviors

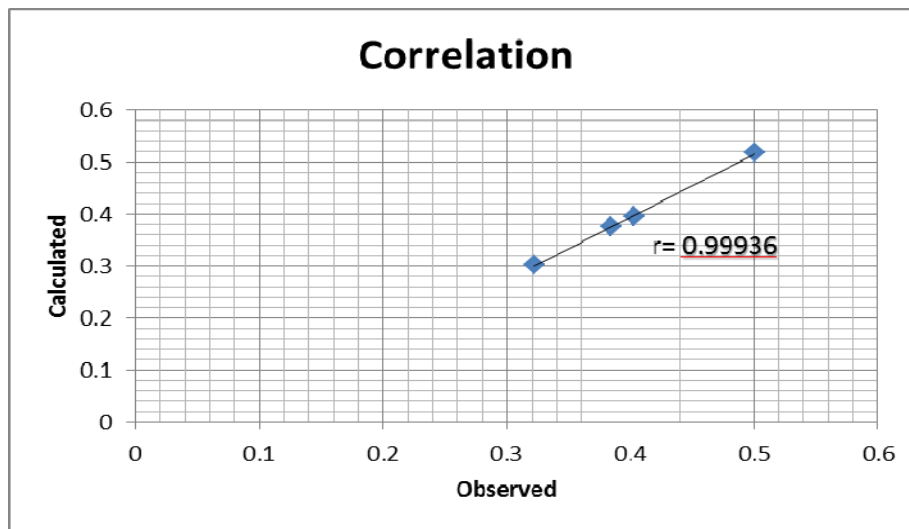


Fig. 4: Correlation between observed and calculated value

The correlation coefficient between observed and mathematically modeled attacking probability (AP) for the observation system was calculated to verify model validity. The value is more than 0.999 (see Figure 4), which indicates the validity of the Markov Chain.

3.2.2 Performance relevance of Manchester United

Figure 5 shows the Performance relevance of passes for Manchester United in the European Champions League Final's first half. The diagram (Transition matrix see Table 1) shows that the passing combinations "Patrice Evra – Park Ji-Sung"; "Michael Carrick - Javier Hernandez" and "Patrice Evra - Javier Hernandez", as well as "Giggs – Park Ji-Sung" were an important contribution and therefore very effective for the team's attack, especially in the front 35m, which is considered the opponent's dangerous area. On the other hand, the passing combinations "Patrice Evra – Michael Carrick"; "Javier Hernandez – Michael Carrick" and "Patrice Evra - Vidic" had a negative impact on the attack in the front 35m, hence reducing the mistake rate

among them would contribute much more to the team's attack. Furthermore, both statistical and mathematical simulation results showed that although Carrick did not pass very often to Hernandez, his passes to him contributed more than his other passes to create an advantageous situation.

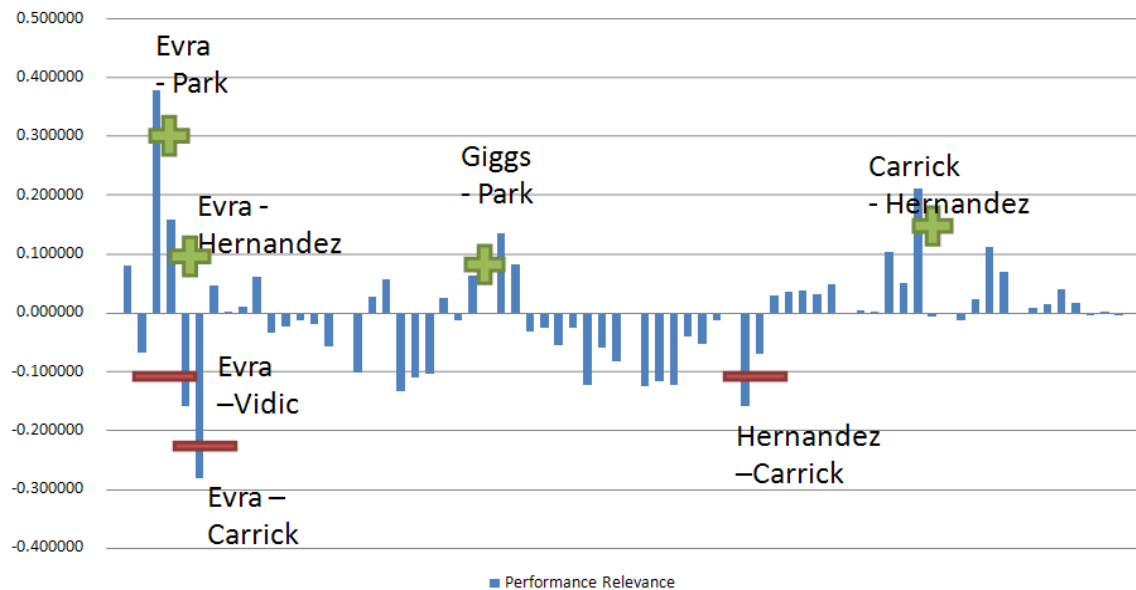


Fig. 5: Performance relevance of passes for Manchester United in the first half

In addition, the table also states that dribbling in the first half was not very important in the team's attacking, however, passes were more effective and efficient.

Table 2: Transition matrix of Manchester United passes in the second half

	Van Der Sar 1	Patrice Evra 3	Rio Ferdinand 5	Wayne Rooney 10	Ryan Giggs 11	Park Ji-Sung 13	Javier Hernandez 14	Nemanja Vidic 15	Michael Carrick 16	Nani 17	Scholes 18	Fabio 20	Antonio Valencia 25	into 35m 35
Van Der Sar 1	10	0	10	25	0	0	5	10	5	5	0	5	5	0
Patrice Evra 3	0	23.33333	0	6.66667	16.66667	3.33333	0	6.66667	0	0	13.33333	0	0	13.33333
Rio Ferdinand 5	6.89655	0	17.24138	10.34483	0	10.34483	0	10.34483	6.89655	3.44828	13.7931	0	13.7931	0
Wayne Rooney 10	0	2.43902	0	26.82927	7.31707	9.7561	4.87805	2.43902	4.87805	4.87805	2.43902	0	9.7561	17.07317
Ryan Giggs 11	0	18.19182	0	22.72727	18.18182	4.54545	4.54545	4.54545	0	0	0	0	4.54545	9.09091
Park Ji-Sung 13	0	0	0	10.52632	10.52632	15.78947	5.26316	15.78947	15.78947	10.52632	0	0	5.26316	0
Javier Hernandez 14	0	0	0	10	20	0	0	0	10	0	10	0	0	50
Nemanja Vidic 15	16.67667	8.33333	20.83333	4.16667	0	4.16667	0	16.66667	4.16667	0	8.33333	0	4.16667	0
Michael Carrick 16	5.55556	5.55556	5.55556	11.11111	16.66667	0	16.66667	0	11.11111	5.55556	0	0	5.55556	5.55556
Nani 17	0	4.34783	0	4.34783	0	4.34783	4.34783	0	4.34783	39.13043	4.34783	0	4.34783	13.04348
Scholes 18	6.25	6.25	31.25	6.25	0	12.5	0	18.75	0	0	0	0	6.25	6.25
Fabio 20	0	0	50	0	0	0	0	0	50	0	0	0	0	0
Antonio Valencia 25	3.7037	0	11.11111	7.40741	0	0	0	3.7037	0	14.81481	11.11111	0	29.62963	0
into 35m 35	0	0	0	0	0	0	0	0	0	0	0	0	0	100
out of 35m 36	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 2 shows the transition matrix of Manchester United passes in the second half, and figure 6 suggests the performance relevance of passes for Manchester United in the second half. The diagram shows that the passes between “Van Der Sar - Rooney”, “Van der Sar - Hernandez”, “Rooney - Hernandez” and “Carrick - Hernandez” ranked at the top, therefore affecting the team's attack into the front 35m areas very positively in the second half. However, the passes between “Rooney - Valencia”, “Evra – Vidic”, “Rooney – Park Ji-sung” and “Hernandez - Giggs” negatively affected the team, which suggests reducing this mistake rate would help setting up the team's offense in the opponent's dangerous areas.

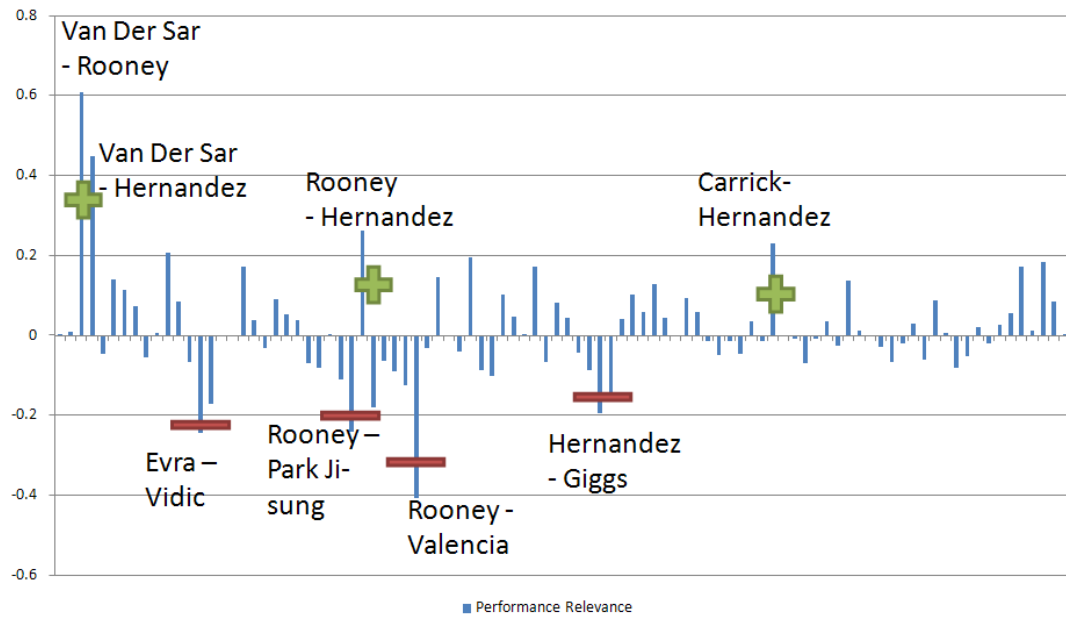


Fig. 6: Performance Relevance of passes for Manchester United in the second half

3.2.3 Performance relevance of FC Barcelona

Table 3: Transition matrix of FC Barcelona passes in the first half

	Valdes 1	Daniel Alves 2	Gerard Pique 3	XAVI Hernandez 6	David Villa 7	Andres Iniesta 8	L Messi 10	Javier Mascherano 14	Sergio Busquets 16	Pedro Rodriguez 17	Eric Abidal 22	Into 35m 35
Valdes 1	7.14286	7.14286	14.28571	0.00000	7.14286	7.14286	0.00000	28.57143	14.28571	0.00000	0.00000	0.00000
Daniel Alves 2	0.00000	16.07143	3.57143	17.85714	5.35714	10.71429	21.42857	3.57143	3.57143	0.00000	0.00000	5.35714
Gerard Pique 3	6.07061	3.03030	15.15152	24.24242	3.03030	6.06061	6.06061	12.12121	15.15152	0.00000	9.09091	0.00000
XAVI Hernandez 6	1.01010	9.09091	0.00000	19.19192	7.07071	20.20202	9.09091	11.11111	6.06061	7.07071	5.05051	3.03030
David Villa 7	0.00000	16.00000	0.00000	12.00000	20.00000	8.00000	12.00000	0.00000	0.00000	4.00000	0.00000	28.00000
Andres Iniesta 8	0.00000	8.97436	2.56410	19.23077	2.56410	12.82051	23.07692	1.28205	8.97436	5.12821	5.12821	6.41026
L Messi 10	0.00000	8.04598	0.00000	14.94253	4.59770	13.79310	28.73563	2.29885	6.89655	4.59770	1.14943	8.04598
Javier Mascherano 14	9.52381	11.90476	9.52381	21.42857	0.00000	2.38095	11.90476	21.42857	4.76190	0.00000	4.76190	0.00000
Sergio Busquets 16	0.00000	6.52174	8.69565	17.39130	0.00000	17.39130	10.86957	4.34783	13.04348	2.17391	13.04348	2.17391
Pedro Rodriguez 17	0.00000	0.00000	0.00000	7.14286	0.00000	25.00000	10.71429	0.00000	3.57143	17.85714	7.14286	17.85714
Eric Abidal 22	0.00000	0.00000	5.41541	10.81081	0.00000	21.62162	8.10811	2.70270	10.81081	13.51351	13.51351	0.00000

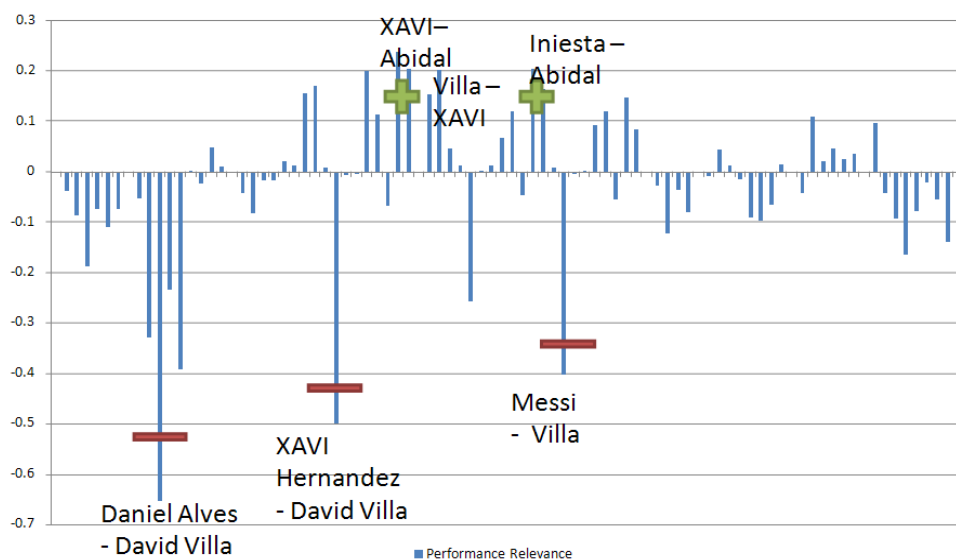


Fig. 7: Performance Relevance of passes for FC Barcelona during the first half

Table 3 shows the transition matrix of FC Barcelona passes in the first half, and figure 7 indicates the performance relevance of passes for FC Barcelona in the first half. From the chart one can see that the rising success rate of the pass combinations “XAVI Hernandez – Eric Abidal”, “David Villa – XAVI Hernandez” and “Andres Iniesta – Eric Abidal” did contribute to the team’s attack of the front 35m area. In addition, reducing the mistake rate of passes between "Daniel Alves - David Villa", "XAVI Hernandez - David Villa", "Messi - David Villa" could further enhance the performance relevance of attacking the 35m area in the first half. Since the majority of these players are playing on the right side, it can be concluded that by reducing the mistake rate on that side, FC Barcelona would make its own attack more effective.

Table 4: Transition matrix of FC Barcelona passes in the second half

	Valdes 1	Daniel Alves 2	Gerard Pique 3	Puyol 5	XAVI Hernandez 6	David Villa 7	Andres Iniesta 8	L Messi 10	Javier Mascherano 14	Keita 15	Sergio Busquets 16	Pedro Rodriguez 17	Eric Abidal 22	Into 35m 35
Valdes 1	0	21.4286	7.14286	0	0	7.14286	0	0	7.14286	7.14286	7.14286	14.2857	0	0
Daniel Alves 2	0	17.3177	0	1.92308	17.3077	11.5385	0	15.3846	7.69231	0	13.4615	0	0	5.76923
Gerard Pique 3	7.40741	3.7037	29.6296	0	25.9259	0	7.40741	0	0	0	3.7037	7.40741	11.1111	0
Puyol 5	0	20	0	0	40	0	0	0	0	0	0	20	0	0
XAVI Hernandez 6	1.09696	14.1304	5.43478	0	16.3044	6.52174	15.2174	15.2174	1.08696	0	6.52174	4.34783	2.17391	5.43478
David Villa 7	0	11.5385	0	0	11.5385	15.3846	3.84615	15.3846	11.5385	0	3.84615	0	0	7.69231
Andres Iniesta 8	0	2.8269	2.8169	1.40845	19.7183	0	22.5352	19.7183	4.22535	0	7.04225	7.04225	4.22535	5.6338
L Messi 10	0	6.09756	2.43902	0	21.9512	0	14.6342	31.7073	1.21951	1.21951	4.87805	1.21951	0	9.7561
Javier Mascherano 14	4.16667	8.33333	4.16667	0	4.16667	20.8333	0	8.33333	16.6667	0	8.33333	0	4.16667	4.16667
Keita 15	0	0	0	0	0	0	25	0	0	0	25	0	0	0
Sergio Busquets 16	0	6.25	6.25	0	8.33333	4.16667	22.9167	14.5833	2.08333	4.16667	18.75	4.16667	6.25	0
Pedro Rodriguez 17	0	0	0	6.89655	3.44828	0	20.6897	6.89655	0	0	6.89655	20.6897	20.6897	6.89655
Eric Abidal 22	6.89655	0	6.89655	0	20.6897	0	17.2414	0	0	0	10.3448	13.7931	13.7931	0

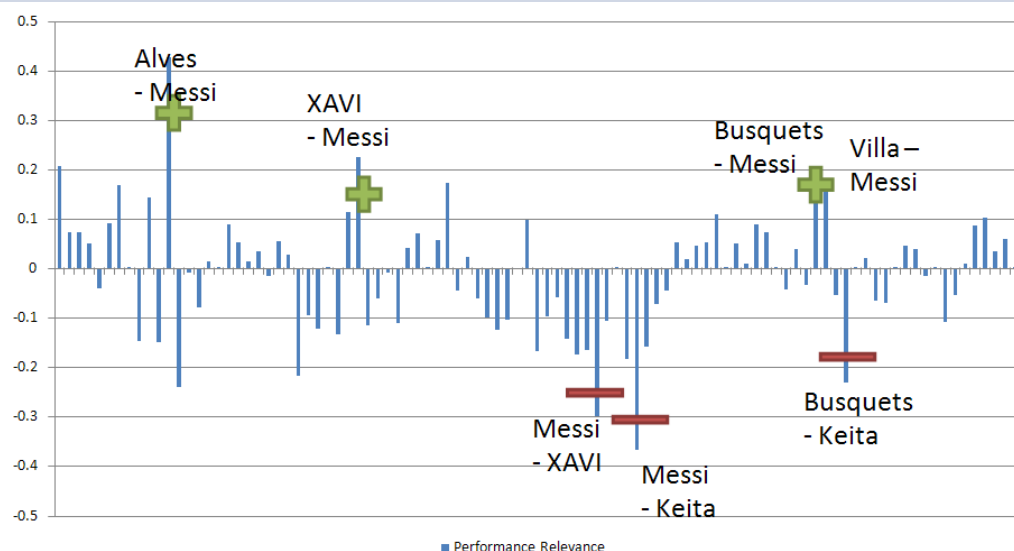


Fig. 8: Performance Relevance of passes for FC Barcelona during the second half

Table 4 shows the transition matrix of FC Barcelona passes in the second half, and figure 8 indicates the performance relevance of passes for FC Barcelona in the second half. The diagram indicates that an improving success rate of the passes between "Daniel Alves - Messi", "XAVI Hernandez - Messi", "Sergio Busquets - Messi" and "David Villa – Messi" would produce better results in the attacking of the front 35 m. However, the pass combinations "Messi - Keita", "Messi - XAVI Hernandez", "Sergio Busquets - Keita" negatively affected the team’s attack of the front 35m area. To reduce the mistake rate of their passes would

make it more efficient in the opponent's dangerous areas. The data suggested that Messi became the playmaker for FC Barcelona in the second half. Furthermore, "Messi's passes to Keita" and "Sergio Busquets' passes to Keita" played a real negative role in the second half, which was probably because of Keita's late substitution in the game.

4. Conclusion

In this study, the Markov model was built to calculate performance relevance. The data clearly indicated that the star players of each team Messi (FC Barcelona) and Rooney (Manchester United) looked to play a more important role in the second half, which was supported by the performance relevance results that were relatively high in relation to them.

5. Reference

- [1] Kuhn, Werner, and Schmidt, Werner. *Analyse und Beobachtung in Training und Wettkampf: Beitrage und Analysen zum Fussballsport IV*. Sankt Augustin: Academia Verlag, 1991.
- [2] J. Bangsbo, T. Reilly, and C. Hughes. *Science and football*. Spon Press, 1997.
- [3] H.Zhang, and A. Hohmann. Performance Diagnosis through Mathematical Simulation in Table Tennis Game. *Journal of Shanghai University of Sport*. 2004, **28**(2): 68-72.
- [4] N. Hirotsu, and M. Wright. An evaluation of characteristics of teams in association football by using a Markov process model. *The Statistician*. 2003, **52**: Part 4, pp. 591-602.
- [5] N. Hirotsu, and M. Wright, Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions. *Journal of the Operational Research Society*. 2002, **53**: 88-96.
- [6] M. Pfeiffer, A. Hohmann. and M. Buehrer. Computersimulation zur Bestimmung der Leistungsrelevanz taktischer Verhaltensweisen bei der FIFA WM 2006.5. Dvs-sportspiel-Symposium 2006, Universitaet Flensburg.
- [7] C. Robson. *Real world research*. Oxford: Blackwell Publishers, 2002.
- [8] Pfeiffer, M., Zhang, H and Hohmann, Andreas. A Markov Chain Model of Elite Table Tennis Competition. *International Journal of Sport Science and Coaching*. 2008, **5**(2): 205-222.
- [9] M. Lames. Leistungsdiagnostik durch Computersimulation. Frankfurt/Main: Harri Deutsch. 1991, pp.7-257.
- [10] M. Pfeiffer. *Leistungsdiagnostik im Nachwuchstraining der Sportspiele*. Koeln: Sportbuch Strauss, 2005.
- [11] W. Greve. and D. Wentura. Wissenschaftliche Beobachtung: Eine Einfuehrung, PVU/Beltz, Weinheim 1997.
- [12] J. R. Landis, and G. G. Koch. The measurement of observer agreement for categorical data. *In: Biometrics*. 1977, **33**: 159-174.

Forschung für den Leistungssport

IAT – Postfach 100841 – 04008 Leipzig

Prof. Dr. Arndt Pfützner
Direktor

Tel.: 0341 4945 100
Fax: 0341 4945 400
iat@iat.uni-leipzig.de

Tianbiao Liu
Lehrstuhl Sport I
Institut für Sportwissenschaft
Universitätsstr. 30
95440 Bayreuth

Leipzig, 15.01.2014

Sehr geehrter Herr Liu,

vielen Dank für die Einsendung des Beitrags

Liu, T., Castro, C., Sörgel, V. & Hohmann, A.
Systematische Spielanalyse im Frauenfußball
Finale des Algarve Cups 2012 Deutschland – Japan

zur Veröffentlichung in unserer Zeitschrift für angewandte Trainingswissenschaft.

Ich freue mich Ihnen mitteilen zu können, dass das Redaktionskollegium einer Veröffentlichung des Beitrags zugestimmt hat. Die Publikation ist für das Jahr 2015 vorgesehen.

Mit freundlichem Gruß



Prof. Dr. Arndt Pfützner
Direktor IAT

Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

Ein Institut im Verein IAT/FES e.V. | Amtsgericht Leipzig VR 1227 | UST-ID: DE 141510029 | ST.-Nr.: 231/140/04294
Marschnerstraße 29 | 04109 Leipzig | Tel.: 0341 4945 - 01 | Fax: 0341 4945 - 400 | Mail: iat@iat.uni-leipzig.de | www.sport-iat.de
Commerzbank | BLZ: 860 800 00 | Konto: 125 675 000 | IBAN: DE15 86080000 0125675000 | Swift-BIC: DRESDEFF860
Deutsche Kreditbank | BLZ: 120 300 00 | Konto: 100 5430 028 | IBAN: DE45 12030000 1005430028 | Swift-BIC: BYLADEM1001

Liu,T, Castro,C, Sörgel,V. & Hohmann, A.¹

Systematische Spielanalyse im Frauenfußball

Finale des Algarve Cups 2012 Deutschland – Japan

Systematic analysis of women's football game

Algarve Cup 2012 final Germany – Japan

Korrespondenzadresse

Tianbiao Liu
Univerisität Bayreuth
Institut für Sportwissenschaft
Lehrstuhl Sportwissenschaft I
D-95440 Bayreuth
LTBVANE@Gmail.com

¹ Universität Bayreuth – Institiut für Sportwissenschaften

Systematische Spielanalyse im Frauenfußball

Finale des Algarve Cups 2012 Deutschland – Japan

Zusammenfassung

Im Rahmen einer Systematischen Sportspielbeobachtung im Frauenfußball wurde das Finale des Algarve Cup 2012, Deutschland gegen Japan (Endstand 4:3), beobachtet und analysiert. Auf der Basis der Videoaufzeichnung des Spiels wurden sämtliche Pässe innerhalb des Spiels sowohl in Bezug auf die räumlichen Spielfeldzonen als auch in Bezug auf die beteiligten Spieler mit Hilfe der beiden Auswertungsmethoden Data mining und Markov-Kette vertieft ausgewertet. Im Ergebnis der Leistungsdiagnostik konnten über die deskriptiven Merkmale des Spielverlaufs, wie z.B. Länge der Ballkontrollketten, räumliche Ballverteilung auf dem Spielfeld (heat maps) und mannschaftlicher Angriffsindex, hinaus, auch tiefer liegende, ergebnisrelevante Muster des mannschaftlichen und räumlichen Passspiels identifiziert werden.

Schlüsselwörter: Fußball, Algarve Cup 2012, Sportspielbeobachtung, Spielanalyse, Data mining, Markov-Kette

Systematic analysis of a women's football match

Algarve Cup 2012 final: Germany vs. Japan

Abstract

Using a systematic method of sports game analysis, the author of this paper has observed and analyzed the final match of the 2012 Algarve Cup, played out between Germany and Japan (final result: 4:3). Based on the match footage, the author recorded game zones and player numbers, before using data mining and Markov chain methods to analyze all passes. The results of this study produced a description of the game, including the length of ball-controlling chains, distribution of the ball on the ground (heat maps) and a team attacking index. In addition, the study identifies the playing patterns based on player modeling and zone modeling, offering a deeper understanding of the game.

Keywords: football, Algarve Cup 2012, sports game observing, game analysis, data mining, Markov chain

Einleitung

Der internationale Frauenfußball hat sich in den letzten zwei Jahrzehnten in vielerlei Hinsicht rasant entwickelt. So besuchten bei der Weltmeisterschaft 2011 in Deutschland durchschnittlich 26.428 Zuschauer und insgesamt 2.575 akkreditierte Journalisten die Spiele vor Ort und weltweit verfolgten ca. 63 Millionen Zuschauer das Turnier vor den Fernsehschirmen. Die gesteigerte Bedeutung des Frauenfußballs wird jedoch nicht alleine durch die Medienattraktivität der Weltmeisterschaft belegt. Vielmehr spiegelt sich die positive Leistungsentwicklung auch im deutlich verringerten Leistungsabstand zwischen den Topnationen und deren Verfolgern wider (FIFA, 2011), weshalb die Zahl der Qualifikationsspiele auf 355 und damit gegenüber der WM 1991 auf mehr als das Dreifache angestiegen ist. Gleichzeitig hat die spielerische und technisch-taktische Qualität des Frauenfußballs erheblich zugenommen und bedarf daher einer ständigen sportwissenschaftlichen Analyse, um im Wettbewerb der Weltspitze bestehen zu können (AFC, 2011; FWF, 2011).

Der Algarve-Cup ist neben dem Olympischen-Fußballturnier sowie der Welt- und Europameisterschaft das weitaus wichtigste Frauenfußballturnier für Nationalmannschaften. Da das Turnier immer zu Jahresbeginn stattfindet, eignet sich dieses Aufeinandertreffen der weltbesten Frauentteams in ausgezeichneter Weise für eine frühzeitige spielerische und technisch-taktische Standortbestimmung der führenden Nationen. Im Algarve-Cup 2012 hat Deutschland zum zweiten Mal nach 2006 das Turnier gewonnen, nachdem die Mannschaft im Finale Japan mit 4:3 Toren besiegt hat. Da Japan aktuell den Weltmeister im Frauenfußball stellt, interessiert aus deutscher Sicht die Frage, welche Besonderheiten der untersuchte Spielverlauf aufgewiesen hat und welche spieltaktischen Faktoren zum Erfolg der Deutschen Mannschaft beigetragen haben. Um diese Fragen zu beantworten und Handlungsempfehlungen für zukünftige Spiele geben zu können, muss das individuelle und mannschaftliche Spielverhalten mit geeigneten Verfahren und Methoden der systematischen Sportspielbeobachtung untersucht werden.

Untersuchungsmethodik

Verfahren der systematischen Spielbeobachtung

Im Fußball werden schon seit den 1970er-Jahren vielfältige deskriptive Statistiken zur Beschreibung des Spielverlaufs sowie des individuellen und mannschaftlichen Spielverhaltens angewandt (Abb. 1; zu den einzelnen Methoden Kuhn, Maier & Wiepking, 1977, Kuhn & Maier, 1980; Kuhn, Otto & Steiner, 1992; Lames, 1992; Hohmann & Rommel, 1994; Bangsbo, Reilly & Hughes, 1997; Alcock, 2010). Die Entwicklung neuer Computer- und Datenbanken-Technologien hat die Trainingswissenschaft zunehmend in die Lage versetzt, neben den herkömmlichen beschreibenden Spielstatistiken (wie z.B. Ballbesitz-, Torschuss- und Zweikampfbilanzen) auch komplexe mathematisch-statistische Methoden zur leistungsdiagnostischen **Analyse und Simulation** des Spielverlaufs zu nutzen. Während unter spielanalytischem Aspekt vor allem *Neuronale Netzwerke* (Perl, 2002; Hohmann & Pfeifer, 2010; Memmert & Perl, 2006) angewandt wurden, haben sich zur mathematischen Simulation des Spielverhaltens beispielweise *Markov-Ketten* (Hirotsu & Wright, 2002, 2003; Pfeiffer, Zhang & Hohmann, 2010; Knoessel, 2006)

besonders bewährt.

Systematische Sportspielbeobachtung		
Spielbeschreibung	Spielanalyse	Spielsimulation
Ziel: <u>Spielstatistiken</u> <i>Beispiele:</i> Ballkontrollphasen, „Heat maps“, Angriffs-Index	Ziel: <u>Spielmuster</u> <i>Beispiel:</i> Passfolgen (Data mining)	Ziel: <u>Spielerfolgskfaktoren</u> <i>Beispiel:</i> Spielzüge (Markov-Ketten)
Zugang		Information

Abb. 1: Ziele und Methoden der systematischen Sportspielanalyse

In der vorliegenden Untersuchung des Endspiels um den Algarve Cup 2012 wurden ausschließlich strukturierte Beobachtungsverfahren der systematischen Spielbeobachtung eingesetzt, die halbstandardisierten Methoden des Scoutings wurden ausgeklammert (hierzu ausführlich Lames, 1994). Die Beobachtungen wurden auf der Basis von Videoaufzeichnungen durchgeführt, da auf diese Weise jede Spielaktion präzise erfasst und analysiert werden kann. Das Videomaterial wurde vom Deutschen Fußball-Bund (DFB) zur Verfügung gestellt und die Spielbeobachtung von geschulten Mitarbeitern des Lehrstuhls für Trainings- und Bewegungswissenschaft an der Universität Bayreuth durchgeführt.

Im Rahmen der mathematisch-statistischen Spielanalyse wurde anstelle der mustererkennenden neuronalen Netzwerk-Verfahren die Auswertungsmethode des Data minings (Wang, Yin & Wang, 2009; Pan, 2010; Liu, 2011) verwendet. Diese Methode hat sich bereits zur Spielanalyse in den Sportsportarten Tischtennis (Yu, Zhang & Hu, 2008) und Basketball (Bhandari, Colet, Parker, Pines, Pratap & Ramanujam, 1997) bewährt. Im Hinblick auf die weitergehende Zielstellung der Spielsimulation wird bei der vorliegenden leistungsdiagnostischen Studie auf das bewährte Verfahren der Markov-Kette zurückgegriffen.

Datenerhebung

Die Datenerhebung beschränkt sich ausschließlich auf Aktionen der ballbesitzenden Spielerinnen und registriert dabei sowohl (1.) jede einzelne *Spielerin*, die an den Passfolgen der ballbesitzenden Mannschaft beteiligt ist, als auch (2.) jede räumliche *Spielfeldzone*, aus der der Ball aus- und eintritt. Neben der Erfassung aller am Spiel beteiligten japanischen und deutschen Stamm- und Wechselspielerinnen wurde das Spielfeld in 30 Zonen aufgeteilt (Abb.2). Die offensiven, im Mittelfeld gelegenen und defensiven Spielfeldzonen unterliegen in der zweiten Halbzeit dem Wechsel der Spielrichtung der Mannschaften und werden deshalb zahlenmäßig so umcodiert, dass die Nummerierung nach wie vor der Spielrichtung entspricht.



Abb. 2: Einteilung und Nummerierung der räumlichen Spielfeldzonen ($n = 30$)

Über die Spielerinnennummer und Spielfeldzonen hinaus wurden alle technischen und taktischen Aktionen durch die Vergabe eines Zahlenindex vorab definiert und bei der Beobachtung entsprechend erfasst und codiert (Tab. 1).

Tab. 1 Strukturierung und Codierschema der technisch-taktischen Spielaktionen im Fußball

Index	Sequenzbeginn	Index	Sequenzende	Index	Aktion	Index	Aktion
100	Anstoß	111	Torerfolg	11	Dribbling	21	Lückepass zwischen zwei Verteidigern hindurch
101	Freistoß	112	Torschuss, vom Torwart gehalten	12	Dribbling und abgewehrter Ball	22	Kurzpasse
102	Einwurf	113	Abpraller vom Tor oder Torwart	13	Passen und Starten	23	Doppelpass
103	Eckstoß	114	Fehlschuss (Ausball)	14	Volleypass	24	Hinterlaufen
104	Torschuss	116	Ausball/Einwurf	15	Ballannahme	25	Langer Querpass (Flügelwechsel)
105	Angriffsbeginn	117	Ausball/Abstoß	16	Torschuss	26	Steilpass
106	Ballgewinn des Torwarts	118	Ausball/Eckball			27	Solodribbling
107	Ballgewinn durch Schiedsrichter-	119	Abgewehrter Pass/Schuss			28	Überwinden der Abseitsfalle

entscheidung				
	120	Ausball/	29	Flanke/Querpass
	121	Dribbling/ Ballverlust	30	Ballabwehr aus gefährlicher Zone ("Klären")
	122	Abseits		
	123	Verletzung/ Spielunterbrechung		

Bei der Videobeobachtung werden die chronologisch aufeinanderfolgenden technisch-taktischen Aktionen beider Mannschaften mit Hilfe der jeweiligen Indexnummer registriert und in einer Tabelle zeilenweise notiert (Tab.2). Der Gesamtumfang aller Einzelaktionen der beiden Finalteilnehmer betrug $n = 1.400$.

Tab. 2 Exemplarische Darstellung der chronologisch geordneten Datenerfassung

Event NO.	Min	Sec	Half	Angriffs- staffette	Event in der Staffette	SpielerNO.	Nächster Spieler NO.	Zone	Nächste zone	Event	Nächstes Event
1	0	0	1	1	1	13	14	0	18	100	13
2	0	0	1	1	2	14	20	18	18	13	14
3	0	2	1	1	3	20	36	14	119	14	119
4	0	10	1	2	1	23	23	23	23	106	13
5	0	11	1	2	2	23	2	23	22	13	13
6	0	14	1	2	3	2	11	22	16	13	13
7	0	17	1	2	4	11	2	16	16	13	13
8	0	20	1	2	5	2	17	16	16	13	13
9	0	21	1	2	6	17	17	16	17	13	11

Datenauswertung

Im Anschluss an die Beobachtung und Registrierung der interessierenden Spielelemente erfolgte die Verarbeitung der Beobachtungsdaten in Bezug auf die drei Beobachtungsziele mit Hilfe der jeweiligen statistischen Auswertungsmethoden. Den Einstieg in die Datenauswertung bildeten (1.) die Ermittlung der herkömmlichen *deskriptiven Kennziffern* zu den Häufigkeiten der definierten Spielaktionen und deren personelle Aufschlüsselung mit einer zusammengefasste Darstellung in Form der (1a) Länge der Ballstaffetten und der (1b) Höhe des Angriffsindex.

Im Rahmen der (2.) mathematisch-analytischen Auswertungsschritte wurde das Data mining mit Apriori-Algorithmus gewählt und mit Hilfe des Softwaretools Microsoft SQL server Datamining Add-in® durchgeführt. Die analytische und darauf aufbauend simulative Datenverarbeitung erfolgte mit Hilfe der Software „Markovkette“ (Eigenentwicklung der Universität Bayreuth, Institut für Sportwissenschaft) Die grafische Darstellung der räumlichen und interpersonellen Passmuster geschieht auf der Folie der Spielfeldzonen, die in Form von sog. heat maps über die jeweilige Farbintensität eine unterschiedlich hohe Passdichte dokumentieren.

Gütekriterien

Die *Objektivität* der systematischen Spielbeobachtung wurde mit Hilfe von Cohen's Kappa überprüft (Greve & Wentura, 1997; Landis & Koch, 1977). Der Mittelwert liegt für die Beobachtung der ballbesitzenden Spielerinnen bei $k = 0.53$ und der angespielten Spielfeldzonen bei $k = 0.60$, sodass dem Beobachtungsvorgang eine moderate, aber hinreichende Datengüte attestiert werden kann.

Die *Validität* des Spielbeobachtungsmodells wurde über den Vergleich der empirisch beobachteten mit den theoretisch berechneten Werten ermittelt (Lames, 1991). Abb. 2 zeigt das Streudiagramm mit Regressionsgeraden und Korrelationskoeffizienten des Modells auf der Datenbasis von insgesamt acht analysierten Halbzeiten². Die Korrelation beträgt $r = 0.96$ ($p < 0.01$; $n = 8$) und belegt die gute Übereinstimmung der simulativ berechneten Ergebnisse mit den tatsächlich beobachteten Werten.

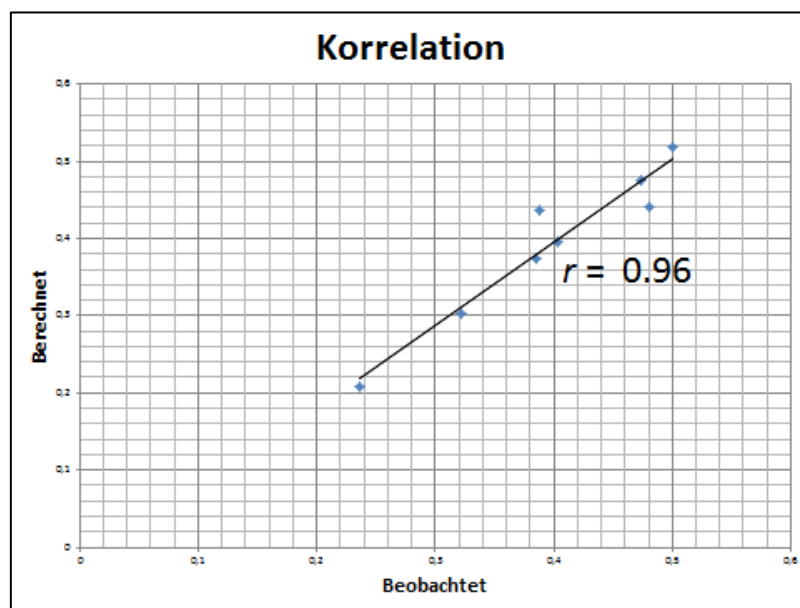


Abb. 2 Korrelation zwischen den theoretisch berechneten und empirisch beobachteten Werten der Spielbeobachtung im Finale des Algarve Cups 2012 sowie der Champions League 2011 (siehe Anmerkung 2)

Untersuchungsergebnisse

(1) Deskriptive Betrachtung des Spielverhaltens

Die durchschnittliche **Länge der Ballkontroll-Staffetten** verweist auf eine besonders in der zweiten Hälfte der Spielzeit zunehmende Überlegenheit der japanischen Frauennationalmannschaft. Die deutsche Mannschaft hat in der ersten Halbzeit $n^{\text{GER}}_{1.\text{Halbzeit}} = 291$ und in der zweiten Halbzeit $n^{\text{GER}}_{2.\text{Halbzeit}} = 163$ Pässe gespielt. Es wurden zuerst $n^{\text{GER}}_{1.\text{Halbzeit}} = 93$

² Zusätzlich zu den vier Werten der beiden Halbzeiten der Endspielpartner (GER und JAP) des Algarve Cups 2012 wurden die vier Werte aus dem parallel untersuchten Champions League-Finale 2011 (FC Barcelona gegen Manchester United) in die Validitätsprüfung einbezogen.

und dann nur noch $n^{GER}_{2.Halbzeit} = 67$ Kontrollsequenzen beobachtet, was sich zugleich auch in einer Verringerung der durchschnittlichen Länge der Ballkontroll-Staffetten von $M^{GER}_{1.Halbzeit} = 4,35$ bzw. $M^{GER}_{2.Halbzeit} = 3,05$ Ballkontakten pro Kontrollsequenz manifestiert. Maximal spielte die deutsche Mannschaft in der ersten Halbzeit $Max^{GER}_{1.Halbzeit} = 13$ Pässe in einer Sequenz und in der zweiten Halbzeit $Max^{GER}_{2.Halbzeit} = 9$ Pässe in einer Passfolge.

Dagegen hat Japan $n^{JAP}_{1.Halbzeit} = 363$ Pässe in der ersten Halbzeit und $n^{JAP}_{2.Halbzeit} = 388$ Pässe in der zweiten Halbzeit gespielt. Dabei wurden $n^{JAP}_{1.Halbzeit} = 89$ und $n^{JAP}_{2.Halbzeit} = 79$ Kontrollsequenzen beobachtet, was durchschnittlich zu $M^{JAP}_{1.Halbzeit} = 5,02$ bzw. $M^{JAP}_{2.Halbzeit} = 5,18$ Pässen pro Kontrollsequenz geführt hat. Als Höchstwert spielte die japanische Mannschaft in der ersten Halbzeit $Max^{JAP}_{1.Halbzeit} = 17$ Ballkontakte in ununterbrochener Folge und in der zweiten Halbzeit erhöhte sich der durchgehende Ballbesitz sogar auf $Max^{JAP}_{2.Halbzeit} = 24$ Ballkontakte (Abb. 4). Der Wilcoxon-Mann-Whitney-U-Test (SPSS, Version 19.0) bestätigt den Befund, dass die durchschnittliche Länge der Ballstaffetten bei der deutschen Mannschaft von der ersten zur zweiten Halbzeit überzufällig abgenommen hat ($Z = -3,01$; $p < 0.01$) und auch gegenüber dem japanischen Team in der zweiten Halbzeit ein systematischer Rückstand bei der Länge der ungestörten Passfolgen eingetreten ist ($Z = -3,80$; $p < 0.01$).

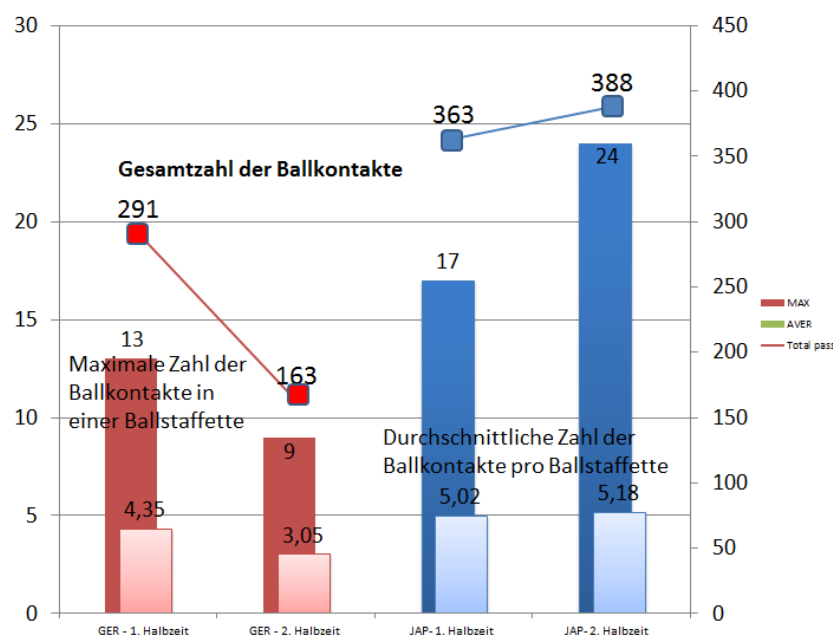


Abb. 2: Länge der Ballkontroll-Staffetten im Finale des Algarve Cups 2012 Deutschland vs Japan

Die mannschaftliche Gesamtleistung wird im Fußball durch die individuellen Leistungen der elf Einzelspielerinnen zwar nicht vollständig determiniert, aber zumindest in hohem Maße beeinflusst (Hohmann & Brack, 1983). Aus diesem Grund ist über die Länge der Ballkontrollstaffetten hinaus auch die Anzahl der Ballkontakte jeder einzelnen Spielerin von Interesse. Neben dieser Bruttokennziffer sind in Abbildung 4 (a) und (b) die Häufigkeiten von individuellen Ballgewinnen und –verlusten dokumentiert. Der Vergleich zwischen der ersten und zweiten Halbzeit zeigt, dass die Spielerinnen der deutschen Mannschaft in der zweiten Hälfte der Spielzeit signifikant weniger Ballkontakte erreichten. Gleichzeitig ist die Anzahl der Balleroberungen deutlich reduziert. Zwar nimmt gleichzeitig auch die Anzahl der Ballverluste in

der zweiten Halbzeit ab, aber die durchschnittliche Ballverlustquote erhöht sich von $M^{\text{GER}}_{1.\text{Halbzeit}} = 20.67\%$ auf $M^{\text{GER}}_{2.\text{Halbzeit}} = 23.96\%$. Diese Verschlechterung der Ballkontrolle geht auf die höheren individuellen Verlustquoten der meisten Spielerinnen im Verlauf des Spiels zurück (Abb. 4 c). Dies hat zur Folge, dass die Einzelspielerstatistik im Vergleich zu Japan (Abb. 4 d) in der zweiten Halbzeit für Deutschland eine signifikant negative Differenz ($\chi^2 = 12.78$; $df = 2$; $p < 0.01$) nachweist.

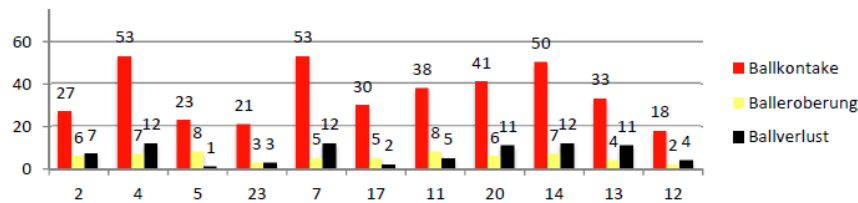


Abb. 4a Ballkontrolle der deutschen Spielerinnen(1. Halbzeit)

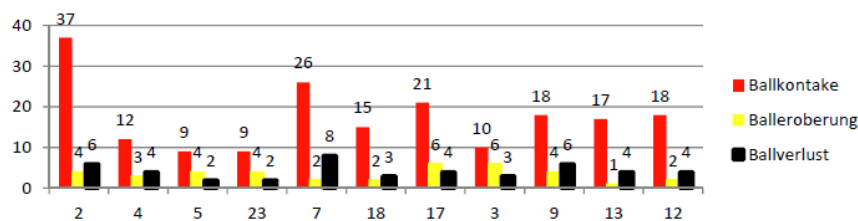


Abb. 4b Ballkontrolle der deutschen Spielerinnen(2. Halbzeit)

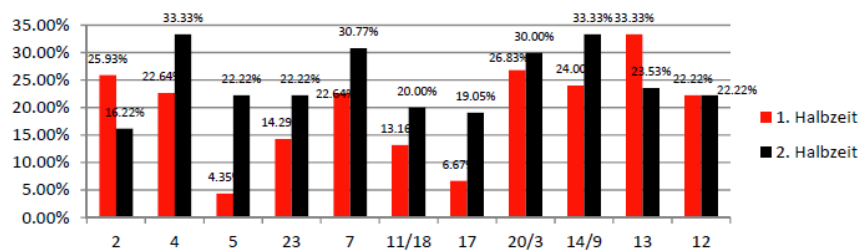


Abb. 4c Ballverlustquoten (Ballverluste pro Ballbesitz) der deutschen Spielerinnen

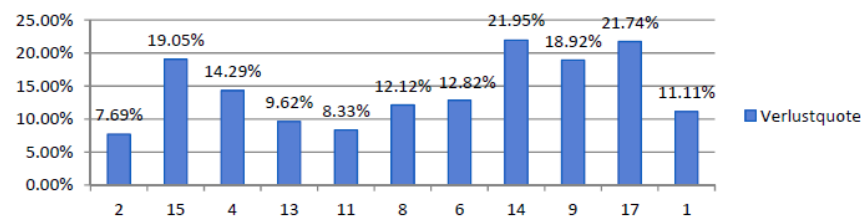


Abb.4d Ballverlustquoten (Ballverluste pro Ballbesitz) der japanischen Spielerinnen (2. Halbzeit)

Abb. 4: Ausgewählte Aspekte der Ballkontrolle durch die deutschen und japanischen Einzelspielerinnen im Finale des Algarve Cups 2012

Neben den Befunden zur Ballkontrolle innerhalb der beiden Teams quantifiziert der **Angriffsindex (AI)** die *Effizienz* eines Angriffs im Verlauf eines Fußballspiels. Um den Angriffsindex zu berechnen, müssen folgende drei Variablen bestimmt werden: die *Spielzeit* (t), die in 5-Minuten-Intervalle aufgeteilt wird, die verschiedenen *Spielaktionen* (n) über deren jeweilige Anzahl innerhalb jedes einzelnen 5-Minuten-Intervalls, sowie die den einzelnen Spielaktionen zugeordneten *Gewichtungspunkte* (x).³ (Liu, 2011). Die Berechnung des Angriffsindex erfolgt anschließend

³ Die in den einzelnen Spielaktionen angewendeten *Techniken* ($x_1 \dots x_n$) wurden wie folgt unterschiedlich

gemäß Formel [1]:

$$[1] \quad AI = f(x, n, t)$$

Abbildung 5 zeigt den AI Deutschlands (GER) im Vergleich zu Japan (JPN) in den beiden Spielzeithälften. Es zeigt sich deutlich, dass Deutschland das Spiel in der 1. Halbzeit überwiegend kontrolliert hat und auch das Halbzeitergebnis von 2:1 Toren unterstreicht die deutsche Dominanz. In der 2. Halbzeit hingegen übernahm Japan zunehmend die Kontrolle des Spielgeschehens und dominierte den Spielverlauf über längere Zeit, auch wenn Japan das Spiel schlussendlich verloren hat. Die Differenz der Werte des Angriffsindex zwischen Deutschland und Japan in der 2. Halbzeit ist statistisch signifikant ($t = -2,30$; $p < 0.05$).

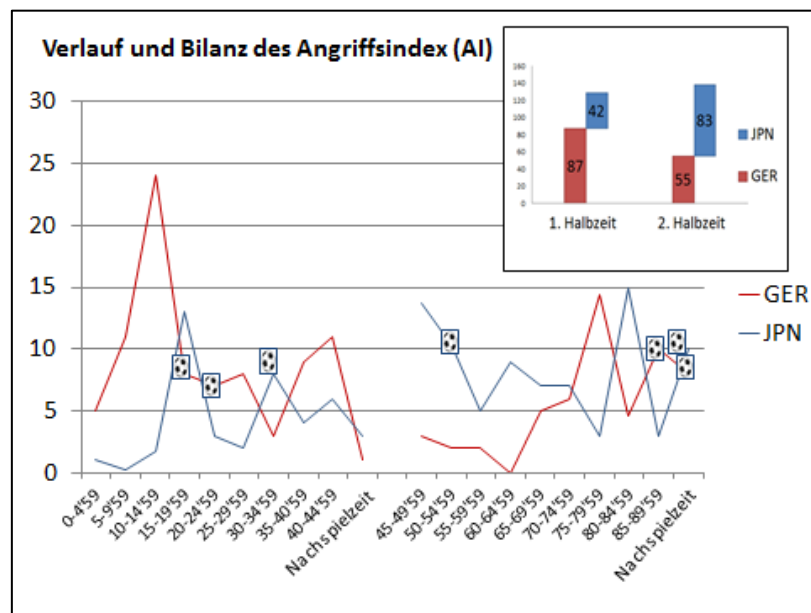


Abb. 5: Zur Ausprägung des Angriffsindex im Finale des Algarve Cups 2012 Deutschland vs Japan

(2) Analytische Betrachtung des Spielverhaltens

Ein Fußballspiel besteht nicht nur aus individuellen Aktionen, sondern auch aus dem Zusammenspiel des gesamten Teams über den zeitlichen Verlauf des Interaktionsprozesses der beiden Spielparteien hinweg. Aus einer entsprechend ganzheitlich-prozessualen Perspektive hat sich bei der Analyse des Spielverlaufs die Methode **Data mining** bewährt. Um die dabei interessierenden Parameter berechnen zu können, wurden sämtliche Passwege innerhalb der beiden Mannschaften analysiert, da die meisten Tore im Fußball nicht zufällig, sondern in Folge eines zielgerichtet funktionierenden Zusammenspiels erzielt werden, das auf diese Weise

gewichtet:

1 Punkt:

- Spielkontrollkette in die 35 Meter Zone
- Spielkontrollkette in den gegnerischen Strafraum
- Ecke/Freistoß in gegnerischen Hälfte/Einwurf in der 35 Meter Zone

2 Punkte:

- Schuss/Freistoß-Schuss/Elfmeter erhalten
- Tor

ermittelt werden kann (vgl. abweichend hiervon Lames, 1992). Gleichzeitig gibt es in jeder Mannschaft bevorzugt aufgesuchte Spielfeldzonen und leistungsbezogen dominante Akteure, sodass beide Informationen zum Spielverhalten unter taktischen Gesichtspunkten relevant sind. Data mining auf der Basis des Apriori Algorithmus – auch „*Frequent Pattern Analysis*“ genannt – stellt bei der zielgerichteten Auswertung großer Datenmengen oder umfangreicher Datenbanken eine weit verbreitete Methode dar. Die Ergebnisse des Data mining beruhen auf den Kriterien „Support“, „Confidence“ bzw. „Probability“ und „Importance“. *Support* bedeutet, dass das gesuchte Muster eine bestimmte Mindesthäufigkeit in den Datensätzen aufweisen muss, um berücksichtigt zu werden. *Confidence/Probability* verdeutlicht die Stärke eines Musters und ergibt sich aus folgende Berechnungsvorschrift [2] (Han & Kamber, 2006):

$$[2] \text{ Confidence/Probability}(X \Rightarrow Y) = P(Y|X) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)}$$

Das dritte Kriterium *Importance*⁴ verleiht solchen Datensätzen, in denen mehr Daten vorkommen, ein höheres Gewicht, weil diese im Rahmen der Auswertung entsprechend eher zu aussagekräftigen Annahmen führen (zum Verfahren des Microsoft MSDN: Ivarsson, 2010). Für die nachfolgende Spielanalyse wurde ein Support von mindestens $n \geq 5$ Zuspielen zwischen zwei Spielern derselben Mannschaft festgelegt und für die Bestimmung eines leistungsrelevanten Musters in den aufgetretenen Passfolgen ein Schwellenwert für die Importance von 45 % gefordert.

(2.1) Räumliche Passmuster im Zonenmodell des Spielfeldes

In Abbildung 6 ist das Zonenmodell mit den dominant aufgetretenen Passmustern der deutschen Mannschaft in der ersten Halbzeit dargestellt.

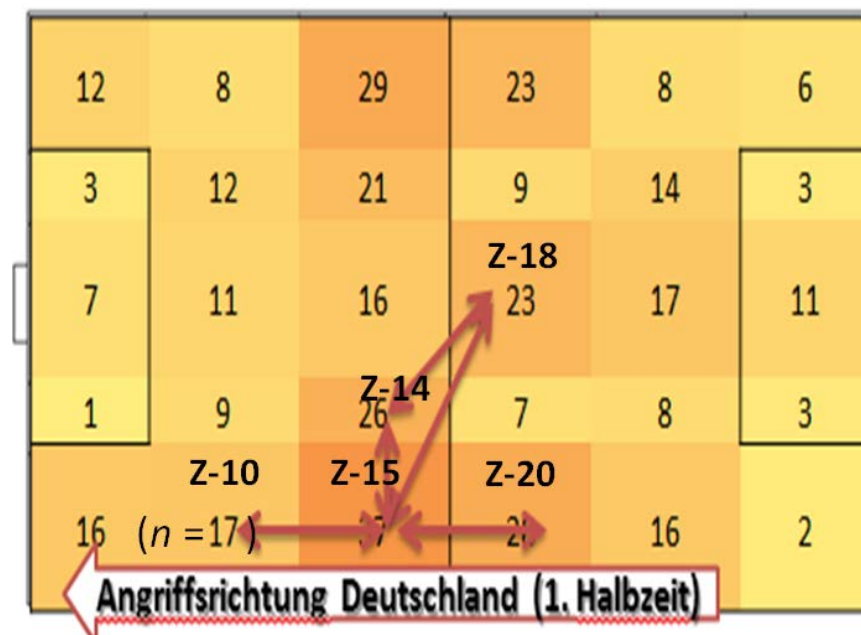


Abb. 6: Dominante räumliche Passmuster der deutschen Mannschaft in der ersten Halbzeit des Endspiels Deutschland vs Japan um den Algarve Cup 2012

⁴ <http://thomasivarssonmalmo.wordpress.com/2010/07/09/calculate-rule-importance-in-ms-association-rules/>

Dabei zeigt sich, dass die Spielfeldareale 10, 14, 15 und 18 tendenziell in den gleichen Spielzügen durchstreift werden. Konkret bedeutet dies, dass beispielsweise bei einem Ballbesitz der deutschen Mannschaft in Zone 15 es ziemlich wahrscheinlich ist, dass die Mannschaft den Ball von dort aus in die Zonen 10, 14 oder 18 weiterleiten wird. Ein solches statistisches Passmuster wird ab einer Importance von 50 Prozent (vgl. Tab. 3) durch die Stärke der grafischen Pfeile ausgewiesen. Insgesamt verdeutlicht die heat map der ersten Halbzeit, dass die Angriffe der deutschen Mannschaft sehr häufig über die linke Seite aufgebaut werden, da sowohl die Probability, als auch die Importance in diesen Zonen sehr hoch ist. Dieses Ergebnis stimmt mit den angegebenen deskriptiven Werten für die Anzahl der dort festgestellten Ballkontakte überein, da die dunkle Färbung dieser Zonen eine hohe Spielintensität anzeigt.

Tab. 3: Ergebnisse des Data minings im räumlichen Zonenmodell der deutschen Mannschaft in der ersten Halbzeit des Endspiels Deutschland vs Japan um den Algarve Cup 2012

Assoziierte Spielfeldzonen (1. Halbzeit)		
Wahrscheinlichkeit (Probability)	Bedeutung (Importance)	Regeln (Association rules)
53 %	0.95	Wenn 15 existiert, existiert gleichzeitig 18
41 %	0.85	Wenn 15 existiert, existiert gleichzeitig 10
50 %	0.73	Wenn 14 existiert, existiert gleichzeitig 18
82 %	0.69	Wenn 18 existiert, existiert gleichzeitig 15
64 %	0.64	Wenn 18 existiert, existiert gleichzeitig 14
78 %	0.59	Wenn 10 existiert, existiert gleichzeitig 15
47 %	0.54	Wenn 15 existiert, existiert gleichzeitig 14

In der zweiten Halbzeit hat sich das Spielverhalten der deutschen Mannschaft deutlich verändert. Das Zonenmodell (Abb. 7) zeigt, dass nun sehr viel aktiver auf der rechten Spielfeldseite gespielt wird. Dass es in der zweiten Spielzeithälfte auf der linken Seite schwieriger geworden war in die gegnerische Hälfte vorzustoßen, wird sowohl durch die Häufigkeiten des Ballbesitzes in der heat map, als auch durch die analytischen Kennziffern des data minings belegt. Die Werte für die Importance weisen beispielsweise für das Passmuster in den Zonen 18, 17 und 11 Folgendes aus: Bei einem Ballbesitz in einer dieser drei aus von dem zentralen Mittelfeld in Richtung der rechten Offensivseite angeordneten Zonen wurden die beiden anderen jeweils zu 100 % in diesem gleichen Spielzug ebenfalls genutzt. Eine mögliche Erklärung für die dadurch entwickelte Verlagerung des deutschen Spiels von der linken auf die rechte Spielfeldseite könnten drei Auswechslungen sein, da alle drei in der zweiten Halbzeit eingewechselten Spielerinnen auf der rechten Seite eingesetzt wurden. Zu beachten ist in diesem Zusammenhang jedoch, dass ungeachtet der guten Qualität der Association Rules die heat map der zweiten Halbzeit die bereits bei der Betrachtung der Gesamtzahl und Länge der mannschaftlichen Ballkontakte festgestellte Negativtendenz belegt, dass die Ballkontrolle in allen Spielfeldzonen, also auch auf der in der zweiten Halbzeit bevorzugten rechten Angriffsseite, im Vergleich zur ersten Halbzeit erheblich abgenommen haben. Das muss in erster Linie wohl auf die verbesserte Spielleistung des Gegners Japan zurückgeführt werden, denn die japanische Mannschaft hat es in der zweiten

Halbzeit geschafft, die deutschen Aktionen wirksam einzuschränken.

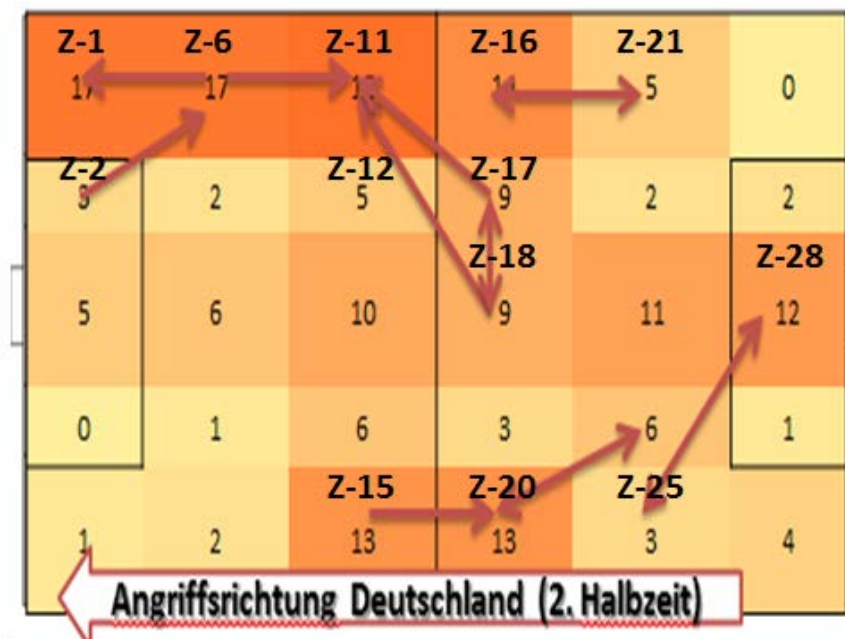


Abb. 7: Dominante räumliche Passmuster der deutschen Mannschaft in der zweiten Halbzeit des Endspiels Deutschland vs Japan um den Algarve Cup 2012

Neben der bevorzugten rechten Spielfeldseite fällt bei der Analyse der dominierenden Passmuster in der zweiten Halbzeit auf, dass zumindest aus der zentralen defensiven Zone (Z-28) das Aufbauspiel der deutschen Mannschaft bevorzugt über das Zusammenspiel mit der linken Spielfeldseite aufrecht erhalten wurde, auch wenn hier meist im Bereich der Mittellinie der Ballvortrag nicht weiter nach vorne fortgesetzt wurde (vgl. Tab. 4).

Tab. 4: Ergebnisse des Data minings im räumlichen Zonenmodell der deutschen Mannschaft in der zweiten Halbzeit des Endspiels Deutschland vs Japan um den Algarve Cup 2012

Assoziierte Spielfeldzonen (2. Halbzeit)		
Wahrscheinlichkeit (Probability)	Bedeutung (Importance)	Regeln (Association rules)
40 %	1.28	Wenn 16 existiert, existiert gleichzeitig 21
100 %	1.25	Wenn 17 und 11 existieren, existiert gleichzeitig 18
50 %	1.03	Wenn 20 existiert, existiert gleichzeitig 24
40 %	0.97	Wenn 17 existiert, existiert gleichzeitig 18
100 %	0.95	Wenn 21 existiert, existiert gleichzeitig 16
100 %	0.95	Wenn 18 und 11 existieren, existiert gleichzeitig 17
75 %	0.88	Wenn 24 existiert, existiert gleichzeitig 20
100 %	0.85	Wenn 25 existiert, existiert gleichzeitig 28
67 %	0.84	Wenn 18 existiert, existiert gleichzeitig 17
100 %	0.77	Wenn 2 existiert, existiert gleichzeitig 6
100 %	0.77	Wenn 18 und 17 existieren, existiert gleichzeitig 11

67 %	0.66	Wenn 18 existiert, existiert gleichzeitig 11
50 %	0.65	Wenn 15 existiert, existiert gleichzeitig 20
38 %	0.52	Wenn 1 existiert, existiert gleichzeitig 11

(2.2) Interpersonelle Passmuster im Spielermodell

Die dominierenden Passmuster zwischen den bestimmten Einzelspielerinnen der deutschen Mannschaft sind für die erste Halbzeit in Abbildung 8 grafisch zusammengestellt.

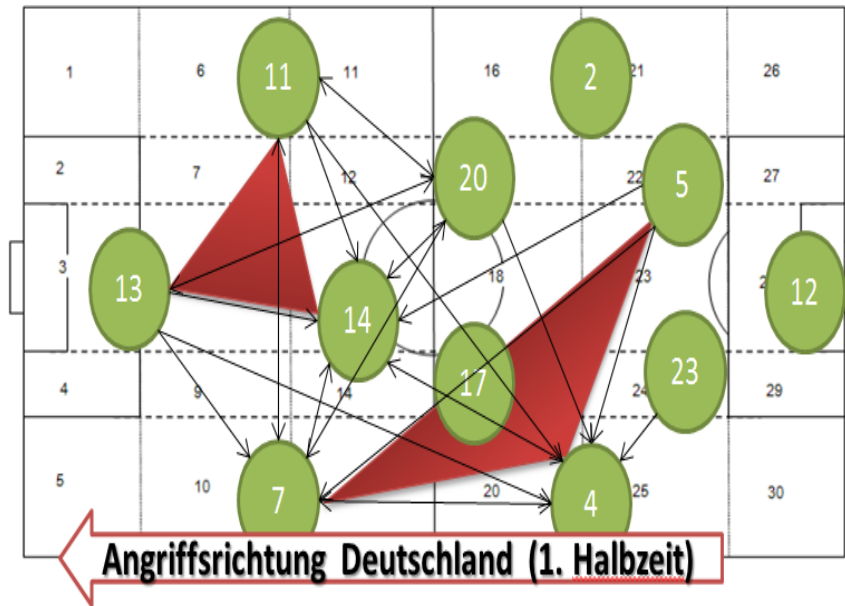


Abb. 8: Dominante interpersonelle Passmuster der deutschen Mannschaft in der ersten Halbzeit des Endspiels Deutschland vs Japan um den Algarve Cup 2012

Unter den vier identifizierten typischen Passmustern (vgl. die Association Rules in Tab. 5) treten zwei Konstellationen aufgrund ihrer hohen Auftretenswahrscheinlichkeit besonders hervor. Diese beiden Spielerkombinationen sind in Abbildung 8 durch rote Dreiecke dargestellt und beruhen auf dem statistischen Befund, dass es unter den verschiedenen beobachteten Passfolgen mit einer ziemlich hohen Wahrscheinlichkeit zu einem Zusammenspiel der drei deutschen Spielerinnen mit den Trikotnummern 7, 4 und 5 sowie den drei Spielerinnen mit den Nummern 11, 14 und 13 kommt. Besonders die letztere Kombinationsform kommt sehr häufig vor und repräsentiert ein für die deutsche Mannschaft sehr typisches offensives Passmuster. Mit Hilfe der Pfeilrichtungen in Abbildung 8 können jedoch nicht nur typische Spielkombinationen, sondern auch bestimmte Schlüsselspieler identifiziert werden. So haben z.B. die deutschen Spielerinnen Nr. 4, Nr. 7, Nr. 14 und Nr. 13 in der ersten Halbzeit eine wichtige Rolle gespielt, wobei die Spielerin Nr. 13 nicht nur als Stürmerin, sondern auch als Spielmacherin im offensiven Vordergrund agiert. Diese Ergebnisse entsprechen insofern den oben (s. Abschnitt 2.1) aufgeführten Befunden, als zumindest drei (Nr. 4, Nr. 7 und Nr. 14) der genannten vier Spielerinnen größtenteils auf der linken Seite gespielt haben.

Tab. 5: Ergebnisse des Data minings zu den interpersonellen Passmustern der deutschen

Mannschaft in der ersten Halbzeit des Endspiels Deutschland vs Japan um den Algarve Cup 2012

Assoziierte Spielerinnen (1. Halbzeit)		
Wahrscheinlichkeit (Probability)	Bedeutung (Importance)	Regeln (Association rules)
55 %	0.79	Wenn 7 und 4 existieren, existiert gleichzeitig 5
83 %	0.59	Wenn 11 und 14 existieren, existiert gleichzeitig 13
38 %	0.56	Wenn 14 und 4 existieren, existiert gleichzeitig 23
38 %	0.56	Wenn 14 und 4 existieren, existiert gleichzeitig 5

Aus Abbildung 9 sowie auch Tabelle 6 kann man entnehmen, dass die Auftretenswahrscheinlichkeiten (Probability) und Bedeutungshöhen (Importance) der meisten Spielerkombinationen gegen über der ersten Halbzeit leicht zurückgegangen sind.

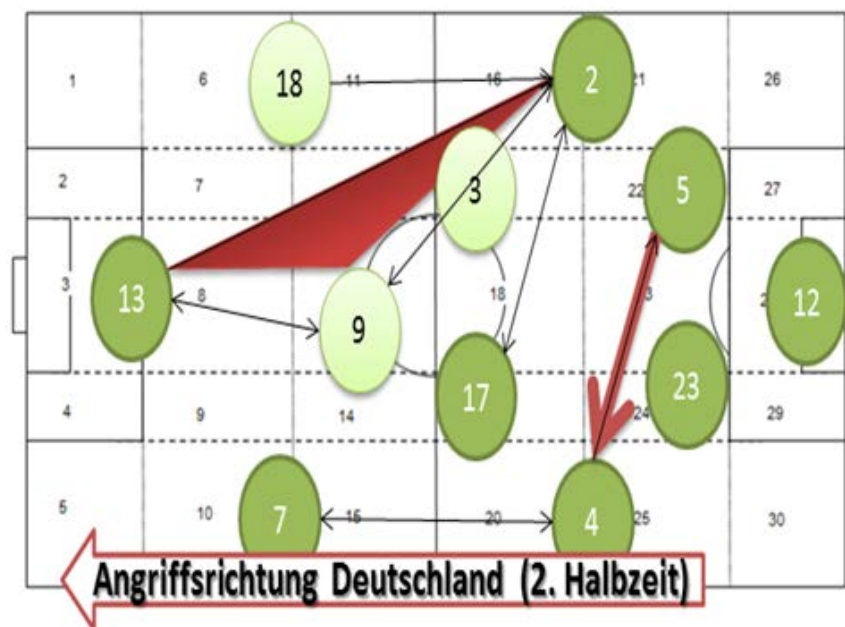


Abb. 9: Dominante interpersonelle Passmuster der deutschen Mannschaft in der zweiten Halbzeit des Endspiels Deutschland vs Japan um den Algarve Cup 2012

Aus Abbildung 9 ist ersichtlich, dass viele Pfeile in Richtung der Mittelfeldspielerin Nr. 2 verlaufen, was belegt, dass diese Spielerin in der zweiten Halbzeit eine wichtige Rolle im deutschen Spiel übernommen hat. Dabei muss jedoch berücksichtigt werden, dass die dazugehörige Assoziationsregel (s. Tab. 6), die ein in die Angriffsmitte gerichtetes Passmuster zwischen den Spielerinnen Nr. 2, Nr. 9 und Nr. 13 beschreibt (rotes Dreieck), zwar eine erhöhte Auftretenswahrscheinlichkeit aufweist, aber seine Bedeutung vergleichsweise gering ausfällt. Ungeachtet dessen hat die deutsche Mannschaft in der zweiten Spielzeithälfte insgesamt große Probleme im taktischen Zusammenspiel. Das Team wurde bei Ballbesitz sehr häufig von den Japanerinnen unterbrochen, sodass sich aus diesem Grund die meisten der in der ersten Halbzeit (aus statistischer Sicht) noch stabil erkennbaren Zuspieldverbindungen (s.o. Abb. 8) zwischen den deutschen Spielerinnen aufgelöst haben. Beispielsweise wurde der vor der Halbzeitpause noch

auffällig kompakte Ballvortrag auf der linken Angriffsseite von der japanischen Mannschaft in der zweiten Halbzeit wirksam unterbunden, sodass die deutsche Spielerin Nr. 7 (ähnlich wie die auf der gegenüberliegenden Angriffsseite eingesetzte Spielerin Nr. 18) den Ball auffallend häufig nur noch zurück und nicht mehr nach vorne abgeben konnte.

Tab. 6: *Ergebnisse des Data minings zu den interpersonellen Passmustern der deutschen Mannschaft in der zweiten Halbzeit des Endspiels Deutschland vs Japan um den Algarve Cup 2012*

Assoziierte Spielerinnen (2. Halbzeit)		
Wahrscheinlichkeit (Probability)	Bedeutung (Importance)	Regeln (Association rules)
44 %	0.66	Wenn 4 existiert, existiert gleichzeitig 5
57 %	0.59	Wenn 5 existiert, existiert gleichzeitig 4
46 %	0.53	Wenn 9 existiert, existiert gleichzeitig 13
80 %	0.50	Wenn 13 und 2 existieren, existiert gleichzeitig 9

(3) Simulative Betrachtung des Spielverhaltens

Um ein Sportspiel als zufallsabhängigen Prozess zu beschreiben und daraufhin verlaufsbezogen und unter besonderer Berücksichtigung der Interaktion zwischen den beiden Spielparteien zu analysieren, hat sich das mathematisch-simulative Vorgehen mit Hilfe der **Markov-Kette** (zur Methode: Lames, 1991; Hughes, 2008) in unterschiedlichen Sportspielen bewährt (z.B. im Tennis: Lames, 1991; Volleyball: Lames & Hohmann, 1997; Tischtennis: Zhang & Hohmann, 2004; Handball: Pfeiffer, 2005; Fußball: Knoessel, 2006; Hughes, 2008). Ganz allgemein gesprochen, ist auch die Methode der Markov-Kette – genauso wie die Methode der Apriori-Algorithmen – den Grundkonzepten des Data Mining⁵ zuzuordnen. Bei der nachfolgenden Prozessanalyse des Fußballspiels werden die verschiedenen, über den individuellen Ballbesitz am Spielfluss beteiligten Spielerinnen als „Zustände“ des Interaktionsprozesses aufgefasst und die beobachteten Übergänge in eine Übergangsmatrix (Tab. 7) überführt. Auf dieser Basis lassen sich die Übergangswahrscheinlichkeiten zwischen den einzelnen Spielzuständen von einem definierten Anfangszustand (Ballgewinn) aus bis hin zu einem definierten, absorbierenden Endzustand (z.B. Tor, Torchance, Eindringen in die 35-m-Zone vor dem gegnerischen Tor oder Ähnliches) berechnen.

Tab. 7: *Übergangsmatrix zwischen den Spielzuständen der deutschen Einzelspielerinnen im Endspiels Deutschland vs Japan um den Algarve Cup 2012*

	Schmidt [#2]	Peter [#4]	Krahn [#5]	Behringer [#7]	Mittag [#11]	Schult [#12]	Mbabi [#13]	Marozsan [#14]	Odebrecht [#17]	Goessling [#20]	Hening [#23]	35-m-Zone
Schmidt	25,9%	0,0%	3,7%	0,0%	7,4%	7,4%	3,7%	3,7%	3,7%	11,1%	3,7%	7,4%
Peter	0,0%	20,8%	0,0%	18,9%	3,9%	0,0%	5,7%	11,3%	1,9%	9,4%	5,7%	13,2%

⁵ "Broadly speaking, Data mining is the process of discovering interesting patterns and knowledge from large amounts of data" (Han & Kamber, 2006, S. 7)

Krahn	0,0%	21,8%	34,8%	8,7%	4,4%	8,7%	0,0%	4,4%	0,0%	4,4%	4,4%	4,4%
Behringer	1,9%	5,7%	0,0%	34,0%	0,0%	1,9%	7,6%	9,4%	7,6%	3,8%	1,9%	11,3%
Mittag	2,6%	10,5%	2,6%	7,9%	42,1%	2,6%	2,6%	2,6%	0,0%	7,9%	2,6%	7,9%
Schult	11,1%	11,1%	5,6%	0,0%	0,0%	22,2%	0,0%	11,1%	0,0%	5,6%	11,1%	0,0%
Mbabi	0,0%	0,0%	0,0%	14,7%	5,9%	0,0%	29,4%	8,8%	0,0%	2,9%	0,0%	26,5%
Marozsan	0,0%	8,0%	0,0%	6,0%	8,0%	0,0%	8,0%	32,0%	4,0%	6,0%	0,0%	18,0%
Odebrecht	3,3%	3,3%	0,0%	0,0%	3,3%	0,0%	10,0%	10,0%	40,0%	6,7%	10,0%	10,0%
Goessling	2,4%	2,4%	0,0%	4,9%	4,9%	0,0%	7,3%	9,8%	7,3%	24,4%	9,8%	9,8%
Hening	9,5%	14,3%	14,3%	9,5%	0,0%	4,7%	0,0%	0,0%	4,8%	14,3%	9,5%	0,0%
35-m-Zone	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%

Mit der Software „Markovkette“ des Instituts für Sportwissenschaft der Universität Bayreuth kann mit Hilfe der Übergangsmatrix die finale Erfolgswahrscheinlichkeit des Übergangs in den interessierenden Endzustand berechnet werden. Um einen taktisch relevanten und zugleich auch häufig(er) erreichten Endzustand zu verwenden, wurde in der vorliegenden Studie das Eindringen in den offensiven Bereich 35 m vor dem gegnerischen Tor (parallel zur Torauslinie) als Angriffsziel definiert. Auf der Basis der originalen Übergangsmatrix kann man in einem nachfolgenden simulativen Arbeitsschritt bestimmte Übergangswahrscheinlichkeiten gezielt erhöhen oder verringern, um die (statistische) Auswirkung auf das Erreichen des Spielziels zu ermitteln. Je höher dabei der Zuwachs an Erfolgswahrscheinlichkeit des Erreichens des definierten Endzustand ausfällt, desto positiver beeinflusst das aus taktischem Interesse vorab manipulierte Passspiel zwischen den betreffenden beiden Spielfeldzonen bzw. Einzelspielerinnen das Erreichen des definierten Endzustands. Auf diese Weise ist man somit in der Lage, Erfolg versprechende „Optimalstrategien“ auszusuchen und virtuell durchzuspielen, die für zukünftige Begegnungen gegen den selben Gegner von Bedeutung sein können (Hohmann & Lames, 2005).

In Tabelle 8 sind die Ergebnisse der ersten Halbzeit zu den mathematisch-simulativ bestimmten Veränderungen bei den Erfolgswahrscheinlichkeiten der deutschen Mannschaft in Bezug auf die ausgewählten Passzuspiele zwischen einzelnen Spielerinnen aufgelistet. Aus den Maximal- und Minimalwerten ergibt sich, dass die Übergänge von Spielerin Nr. 2 zu Spielerin Nr. 13, von Spielerin Nr. 4 zu Spielerin Nr. 13 sowie von Spielerin Nr. 2 zu Spielerin Nr. 14 die am stärksten leistungsverbessernd wirkenden Passfolgen darstellen und eine wichtige Rolle im offensiven Spielprozess spielen könnten. Demgegenüber wirken sich die Übergänge von Spielerin Nr. 4 zu Spielerin Nr. 23 sowie von Spielerin Nr. 4 zu Spielerin Nr. 20 simulativ sehr negativ auf das Angriffsspiel aus. Das bedeutet, dass die virtuell positiv wirkenden Passkombinationen vermehrt angewendet werden sollten und die sich negativ auf das Angriffsverhalten auswirkenden Pässzuspiele eher vermieden oder deren Fehlerraten reduziert werden sollten.

Tab. 8: Mathematisch-simulativ ermittelte Erfolgswahrscheinlichkeiten für das Passzuspiel zwischen den deutschen Spielerinnen in der ersten Halbzeit des Endspiels Deutschland vs Japan um den Algarve Cup 2012

Positive Spielaktionen	Negative Spielaktionen
-------------------------------	-------------------------------

Passzuspiel	Erfolgswahrscheinlichkeit	Passzuspiel	Erfolgswahrscheinlichkeit
Schmidt 2 - Mbabi 13	0.21803	Peter 4 - Hening 23	-0.35978
Peter 4 - Mbabi 13	0.18890	Peter 4 - Goessling 20	-0.21759
Schmidt 2 - Marozsan 14	0.18301	Peter 4 - Schult 12	-0.20345
Schmidt 2 - Odebrecht 17	0.16546	Peter 4 - Behringer 7	-0.19160
Peter 4 - Marozsan 14	0.16303	Peter 4 - Schmidt 2	-0.15212
Behringer 7 - Mbabi 13	0.15525	Mbabi 13 - Behringer 7	-0.14433
Schmidt 2 - Mittag 11	0.14707	Odebrecht 17 - Hening 23	-0.12344
Goessling 20 - Mbabi 13	0.14507	Marozsan 14 - Goessling 20	-0.11238
Schmidt 2 - Goessling 20	0.13535	Behringer 7 - Schult 12	-0.11018
Goessling 20 - Marozsan 14	0.13277	Marozsan 14 - Schult 12	-0.10327
		Goessling 20 - Hening 23	-0.09982

In den beiden Abbildungen 10 und 11 werden vor dem Hintergrund der durch rote Pfeile repräsentierten Befunde aus dem Data mining (s. Abschnitt 2.2) die mit Hilfe der Markovkette ermittelten leistungsrelevanten Positivzuspiele und Negativzuspiele in Form der blauen Pfeile dargestellt. Inhaltlich kann man aus Abbildung 10 darauf schließen, dass in der ersten Halbzeit die Flanken der deutschen Mannschaft von den beiden Außenbahnen des Spielfelds sehr effektiv waren, denn die überwiegende Anzahl der (blauen) Pfeile verläuft in Richtung der beiden zentral spielenden Akteure Nr. 14 und der noch stärker offensiv agierenden Nr. 13. Im Gegensatz zum Data mining bestätigt die simulative Vorgehensweise bei der Spielanalyse eine positive Leistungsrelevanz von nach hinten gespielten Pässen nicht.

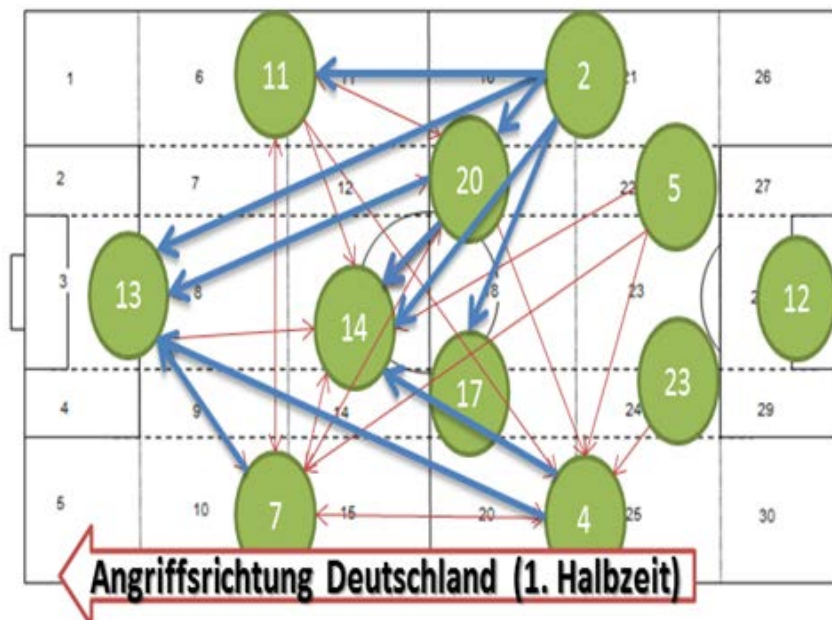


Abb. 10: Mathematisch-simulativ bestimmte, positiv leistungsrelevante Passmuster der deutschen Mannschaft in der ersten Halbzeit des Endspiels Deutschland vs Japan um den Algarve Cup 2012

Auch wenn Rückpässe im Ergebnis des Data minings als auffällig gehäuft auftretende Spielaktionen identifiziert wurden, so wirken sich diese Zuspiele im Rahmen der mathematisch-simulativen Leistungsdiagnostik (vgl. Abb. 11) in der überwiegenden Mehrzahl der Spielerkombinationen negativ auf das Angriffsspiel aus.

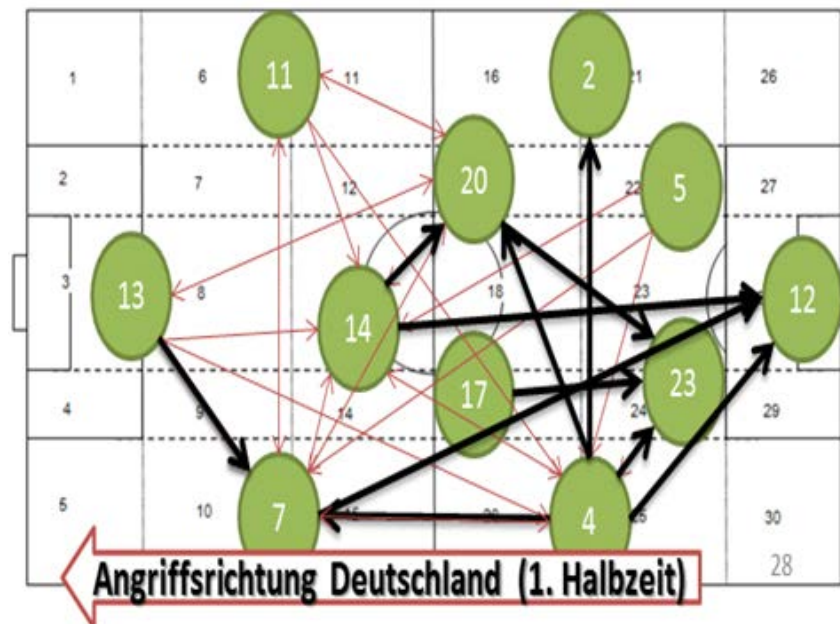


Abb. 11: Mathematisch-simulativ bestimmte, negativ leistungsrelevante Passmuster der deutschen Mannschaft in der ersten Halbzeit des Endspiels Deutschland vs Japan um den Algarve Cup 2012

Die sich auf das Angriffsspiel bzw. das Erreichen der 35-m-Zone vor dem gegnerischen Tor positiv und negativ auswirkenden Passzuspiele der deutschen Mannschaft in der zweiten Halbzeit sind in den Abbildungen 12 und 13 illustriert. Die Pässe von Spielerin Nr. 3 zu Spielerin Nr. 13, von Spielerin Nr. 2 zu Spielerin Nr. 13 sowie von Spielerin Nr. 2 zu Spielerin Nr. 9 wirken sich mathematisch-simulativ am besten aus. Gleichzeitig haben die Pässe von Spielerin Nr. 2 zu Spielerin Nr. 12, von Spielerin Nr. 2 zu Spielerin Nr. 17 sowie von Spielerin Nr. 2 zu Spielerin Nr. 4 einen statistisch negativen Einfluss auf das deutsche Offensivspiel.

Tab. 9: Mathematisch-simulativ ermittelte Erfolgswahrscheinlichkeiten für das Passzuspiel zwischen den deutschen Spielerinnen in der zweiten Halbzeit des Endspiels Deutschland vs Japan um den Algarve Cup 2012

Positive Spielaktionen		Negative Spielaktionen	
Passzuspiel	Erfolgswahrscheinlichkeit	Passzuspiel	Erfolgswahrscheinlichkeit
Bartusiak 3 - Mbabi 13	0.45613	Schmidt 2 - Schult 12	-0.33911

Schmidt 2 - Mbabi 13	0.33302	Schmidt 2 - Odebrecht 17	-0.31854
Schmidt 2 - Popp 9	0.22182	Schmidt 2 - Peter 4	-0.29552
Bartusiak 3 - Schmidt 2	0.17558	Schmidt 2 - Krahn 5	-0.25540
Odebrecht 17 - Mittag 11	0.17271	Schmidt 2 - Bartusiak 3	-0.21381
Odebrecht 17 - Mbabi 13	0.14305	Schmidt 2 - Huth 18	-0.20373
Bartusiak 3 - Behringer 7	0.1399	Behringer 7 - Peter 4	-0.19110
Huth 18 - Popp 9	0.13465	Schmidt 2 - Hening 23	-0.18063
Peter 4 - Popp 9	0.11043	Mbabi 13 - Behringer 7	-0.13729
Peter 4 - Behringer 7	0.10786	Popp 9 - Bartusiak 3	-0.12803

Aus den Abbildungen 12 und 13 kann man entnehmen, dass Spielerin Nr. 2 auf der rechten Spielfeldseite viele positive Kombinationen und Pässe gespielt hat, denen jedoch zugleich eine hohe Anzahl an negativen Passzuspielen gegenüberstehen. Dies ist auf der gegenüberliegenden linken Außenbahn nicht der Fall, wo die defensive Mittelfeldspielerin Nr. 4 durch ausschließlich positiv gekennzeichnete Offensivaktionen auffällt.

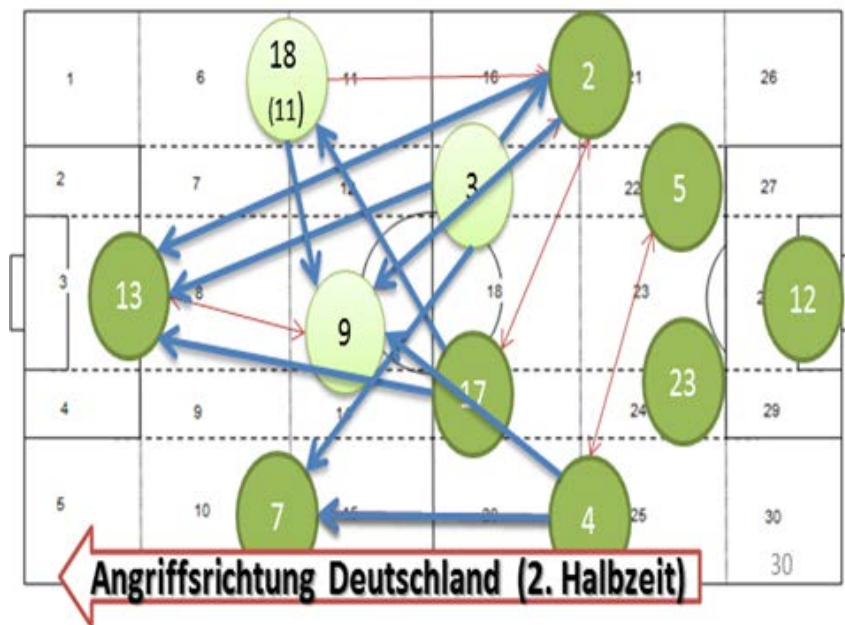


Abb. 12: Mathematisch-simulativ bestimmte, positiv leistungsrelevante Passmuster der deutschen Mannschaft in der zweiten Halbzeit des Endspiels Deutschland vs Japan um den Algarve Cup 2012

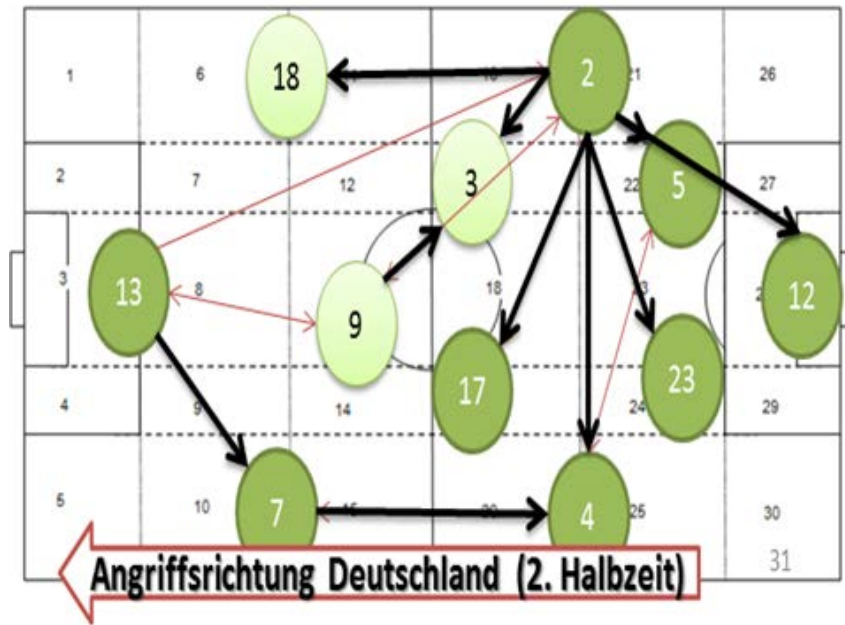


Abb. 13: Mathematisch-simulativ bestimmte, negativ leistungsrelevante Passmuster der deutschen Mannschaft in der zweiten Halbzeit des Endspiels Deutschland vs Japan um den Algarve Cup 2012

Zusammenfassung

Die Analyse des Endspiels um den Algarve Cup 2012 zwischen den Frauennationalmannschaften Deutschlands und des amtierenden Weltmeisters Japan mit Hilfe der fortgeschrittenen Auswertungsmethoden des Data mining hat im Vergleich zu rein deskriptiven Verfahren der Spielbeobachtung zu einem vertieften Verständnis des Spielprozesses beigetragen. Aus inhaltlicher Sicht hat sich in der vorliegenden Studie herausgestellt, dass die deutsche Mannschaft in der ersten Halbzeit des Finales mehr auf der linken und in der zweiten Halbzeit mehr auf der rechten Spielfeldseite agiert hat. Zugleich war das deutsche Offensivspiel mit Ball in der ersten Halbzeit weitaus besser als in der zweiten Halbzeit, was sich sowohl in den beschreibenden, als auch den analytischen Statistiken widerspiegelt hat. Da das deutsche Spiel in der zweiten Halbzeit von den Japanerinnen gut unterbunden wurde, hat sich die Zahl der beobachteten Spielaktionen im Vergleich zur ersten Halbzeit systematisch reduziert. Zusammenfassend haben sich die Methoden des Apriori-Algorithmus und der Markov-Kette als wertvolle Analyseinstrumente auch in dem sehr komplexen Sportspiel Fußball erwiesen.

Danksagung

Der Dank der Autoren gilt dem Deutschen Fußball-Bund (DFB), ohne dessen umfassende Unterstützung die vorliegende Untersuchung nicht möglich gewesen wäre. Dieses

Forschungsprojekt des Erstautors zur Erlangung der Doktorwürde wurde vom Deutschen Akademischen Auslandsdienst (DAAD) finanziell unterstützt.

Literatur

1. AFC (Asian Football Confederation) (2011). *Women's football growing fast*. Zugriff am 08. Dezember 2011 unter <http://www.the-afc.com/en/news/37421-womens-football-grows-fast>.
2. Alcock, A. (2010). Analysis of direct free kicks in the women's football World Cup 2007. *European Journal of Sport Science*, 10 (4), 279-284.
3. Bangsbo, J., Reilly, T., & Hughes, C. (1997). *Science and football*. E&F Spon Press.
4. Bhandari, I., Colet, E., Parker, J., Pines, Z., Pratap, R., & Ramanujam, K. (1997). Advanced scout: Data mining and knowledge discovery in NBA data. *Data Mining and Knowledge Discovery*, 1 (1), 121-125.
5. FIFA (Fédération Internationale de Football Association) (2011). *FIFA Women's World Cup Germany 2011™ Technical Report and Statistics*. Zugriff am 15. April 2013 unter <http://www.fifa.com/mm/document/footballdevelopment/technicalsupport/01/50/87/69/technicalreportfwwcgermany2011.pdf>
6. FWF (Fußball-WM-Frauen) (2011). *Der Frauenfußball entwickelt sich stetig*. Zugriff am 11. November 2011 unter <http://www.fussball-wm-frauen.de/der-frauenfussball-entwickelt-sich-stetig>
7. Greve, W., & Wentura, D. (1997). *Wissenschaftliche Beobachtung: Eine Einführung*. Weinheim: PVU/Beltz.
8. Han, J., Kamber, M., & Pei, J. (2006). *Data mining: concepts and techniques*. San Francisco: Morgan Kaufmann.
9. Hirotsu, N., & Wright, M. (2002). Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions. *Journal of the Operational Research Society*, 53 (1), 88-96.
10. Hirotsu, N., & Wright, M. (2003). An evaluation of characteristics of teams in association football by using a Markov process model. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52 (4), 591-602.
11. Hohmann, A., & Brack, R. (1983). Theoretische Aspekte der Leistungsdiagnostik im Sportspiel. *Leistungssport*, 13 (2), 5-10.
12. Hohmann, A. & Pfeiffer, M. (2010). Applications of Neural Networks in Training Science. *Human Movement Science*, 31 (2), 344-359.
13. Hohmann, A., & Rommel, G. (1994). Spielbeobachtung im Fußball. *Leistungssport*, 22 (6), 41-46.
14. Hohmann, A., & Lames, M. (2005). Trainings- und Wettspielanalyse. In: A. Hohmann, M. Kolb & K. Roth (Hrsg.). *Handbuch Sportspiel* (Beiträge zur Lehre und Forschung im Sport, S. 376–394). Schorndorf: Hofmann.
15. Hughes, M. (2008). Sports Analysis. In: M. Hughes & M. Franks. (Eds.), *The essentials of performance analysis - an Introduction* (pp. 85-97). Chippenham: Routledge.
16. Ivarsson, T. (2010). Calculate Rule Importance in MS Association Rules. Zugriff am 14. April 2013 unter

<http://thomasivarssonmalmo.wordpress.com/2010/07/09/calculate-rule-importance-in-ms-association-rules/>

17. Knoessel, P. (2006). *Systematische Spielbeobachtung bei der FIFA-Weltmeisterschaft 2006™*. Saarbrücken: VDM Verlag Dr. Müller.
18. Kuhn, W., & Maier, W. (1980). Untersuchung zur Identifizierung von Mannschaftsgruppierungen mit ähnlichen Merkmalsstrukturen im Fußball. In R. Andresen & G. Hagedorn (Hrsg.), *Beobachten und Messen im Sportspiel* (S. 161-184). Berlin: Bartels & Wernitz.
19. Kuhn, W., Maier, W. & Lüpsen, H. (1978). *Beiträge zur Analyse des Fussballspiels*. Sankt Augustin: Academia.
20. Kuhn, W., Maier, W. & Wiepking, W. (1977). Untersuchung zur Problematik der Spielzeit im Fußball. Dargestellt an Hand von 24 Spielen der Fussball-Weltmeisterschaft 1974. *Leistungssport*, 7 (1), 22-30.
21. Kuhn, W., Otto, R. & Steiner, J. (1992). Computergestützte Leistungserfassung im Fußball. In W. Kuhn & W. Schmidt (Hrsg.), *Analyse und Beobachtung in Training und Wettkampf. Beiträge und Analysen zum Fußballsport IV* (S. 118-134). Sankt Augustin: Academia.
22. Lames, M. (1991). *Leistungsdiagnostik durch Computersimulation: Ein Beitrag zur Theorie der Sportspiele am Beispiel Tennis*. Frankfurt a. M.: Deutsch.
23. Lames, M. (1992). Probleme von Beobachtungssystemen in den Sportspielen am Beispiel Fußball. In W. Kuhn & W. Schmidt (Hrsg.), *Analyse und Beobachtung in Training und Wettkampf. Beiträge und Analysen zum Fußballsport IV* (S. 135-153). Sankt Augustin: Academia.
24. Lames, M. (1994). *Systematische Spielbeobachtung*. Münster: Philippka.
25. Lames, M., & Hohmann, A. (1997). Zur Leistungsrelevanz von Spielhandlungen im Volleyball. In B. Hoffmann & P. Koch (Hrsg.), *Integrative Aspekte in Theorie und Praxis der Rückschlagspiele* (S. 121-128). Hamburg: Edition Czwalina.
26. Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.
27. Liu, T. (2011). *Applying Data mining to analyze the Attacking Actions between FC Barcelona and Manchester United in European Champions league Finale 2011 (Unveröffentlichte Arbeit)*. Universität Bayreuth: Institut für Sportwissenschaft.
28. Memmert, D. & Perl, J. (2006). Analysis of Game Creativity Development by Means of Continuously Learning Neural Networks. In E. F. Moritz & S. Haake (Eds.), *The Engineering of Sport 6, Vol. 3* (pp. 261–266). New York: Springer.
29. MSDN, Microsoft Association Algorithm Technical Reference. Zugriff am 14. April 2013 unter <http://msdn.microsoft.com/en-us/library/cc280428.aspx>
30. Pan, C. (2010). Appliance of Apriori Algorithm on technical-tactics analysis of football. *Computer Knowledge and Technology*, 31, 66.
31. Perl, J. (2002). Dycon: Ein neuer Ansatz zur Modellierung und Analyse von Sportspiel-Prozessen mit Hilfe neuronaler Netze. In K. Ferger, N. Gissel & J. Schwier (Hrsg.), *Sportspiele erleben, vermitteln, trainieren*, (S. 253-265). Ort: Verlag.
32. Pfeiffer, M. (2005). *Leistungsdiagnostik im Nachwuchstraining der Sportspiele*. Köln: Sport und Buch Strauß.
33. Pfeiffer, M., Hohmann, A. & Bühner, M. (2006). Computersimulation zur Bestimmung der

Leistungsrelevanz taktischer Verhaltensweisen bei der FIFA WM 2006. In M. Raab, A. Arnold, K. Gärtner, J. Köppen, C. Lempertz, N. Tielemann & H. Zastrow (Hrsg.), *Zukunft der Sportspiele: fördern, fordern, forschen* (Tagungsband zum 5. Sportspiel-Symposium) (S. 195-198). Flensburg: University Press.

34. Pfeiffer, M., Zhang, H., & Hohmann, A. (2010). A Markov Chain Model of Elite Table Tennis Competition. *International Journal of Sports Science and Coaching*, 5 (2), 205-222.
35. Wang, B., Yin, Z., & Wang, L. (2009, December). Research of association rules in analyzing technique of football match. *Power Electronics and Intelligent Transportation System (PEITS), 2nd International Conference in Shenzhen*, 3, 178-180. Institute of Electrical and Electronics Engineering (IEEE).
36. Yu, L. J., Zhang, H., & Hu, J. J. (2008). Computer diagnostics for the analysis of table tennis matches. *International Journal of Sports Science and Engineering*, 2 (3), 144-153.
37. Zhang, H., & Hohmann, A. (2004). Performance Diagnosis through Mathematical Simulation in Table tennis Game. *Journal of Shanghai University of Sport*, 28 (2), 68-72.

Die Autoren

Tianbiao LIU M.Sc., wissenschaftlicher Mitarbeiter am Lehrstuhl Trainings- und Bewegungswissenschaft des Instituts für Sportwissenschaft an der Universität Bayreuth. Aktuelle Arbeitsschwerpunkte: Sportspielforschung.

Diplom-Sportwissenschaftler Christian CASTRO, ehemaliger Diplomand am Lehrstuhl Trainings- und Bewegungswissenschaft des Instituts für Sportwissenschaft an der Universität Bayreuth.

Diplom-Sportwissenschaftler Vincent SÖRGEL, ehemaliger Diplomand am Lehrstuhl Trainings- und Bewegungswissenschaft des Instituts für Sportwissenschaft an der Universität Bayreuth.

Prof. Dr. Andreas Hohmann; Lehrstuhlinhaber für Trainings- und Bewegungswissenschaft an der Universität Bayreuth. Aktuelle Arbeitsschwerpunkte: Talentforschung, Trainingsadaptation, Leistungsdiagnostik.